

Heteroskedasticity and Spatiotemporal Dependence

Robust Inference for Linear Panel Models with Fixed Effects*

Min Seong Kim[†]
Department of Economics
Ryerson University

Yixiao Sun[‡]
Department of Economics
UC San Diego

This version: June 2013 (Long Version)

Abstract

This paper studies robust inference for linear panel models with fixed effects in the presence of heteroskedasticity and spatiotemporal dependence of unknown forms. We propose a bivariate kernel covariance estimator that nests existing estimators as special cases. Our estimator improves upon existing estimators in terms of robustness, efficiency, and adaptiveness. For distributional approximations, we considered two types of asymptotics: the increasing-smoothing asymptotics and the fixed-smoothing asymptotics. Under the former asymptotics, the Wald statistic based on our covariance estimator converges to a chi-square distribution. Under the latter asymptotics, the Wald statistic is asymptotically equivalent to a distribution that can be well approximated by an F distribution. Simulation results show that our proposed testing procedure works well in finite samples.

Keywords: Adaptiveness, Panel HAC estimator, F-approximation, Fixed-smoothing asymptotics, Fixed-effects 2SLS, Increasing-smoothing asymptotics, Optimal bandwidth, Spatiotemporal dependence

JEL Classification Number: C13, C14, C23

*We thank Cheng Hsiao, an associate editor and two anonymous referees for helpful suggestions. We also thank Brendan Beare, Alan Bester, Otilia Boldea, Pierre-Andre Chiappori, Tim Conley, Gordon Dahl, Feico Drost, Graham Elliott, Patrik Guggenberger, Jinyong Hahn, James Hamilton, Christian Hansen, Hiroaki Kaido, Ivana Komunjer, Esfandiar Maasoumi, Jan Magnus, Bertrand Melenberg, Leo Michelis, Choon-Geol Moon, Philip Neary, Benoit Perron, Elena Pesavento, Ingmar Prucha, Andres Santos, and Halbert White for valuable comments. Sun gratefully acknowledges partial research support from NSF under Grant No. SES-0752443.

[†]Correspondence to: minseong.kim@ryerson.ca. Department of Economics, Ryerson University, 350 Victoria Street, Toronto, Ontario M5B 2K3, Canada.

[‡]visun@ucsd.edu. Department of Economics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0508, USA.

1 Introduction

This paper studies robust inference for linear panel models with fixed effects in the presence of heteroskedasticity and spatiotemporal dependence of unknown forms. As economic data is potentially heterogeneous and correlated in unknown ways across individuals and time, robust inference in the panel setting is an important issue. See, for example, Bertrand, Duflo and Mullainathan (2004) and Petersen (2009). The main interest in this problem lies in (i) how to construct covariance estimators that take the correlation structure into account; (ii) how to approximate the sampling distribution of the associated test statistic; and (iii) how to select smoothing parameters in finite samples.

Regarding covariance estimation, we propose a bivariate kernel estimator. In order to utilize the kernel in the spatial dimension, we need *a priori* knowledge about the dependence structure. It is often assumed that the covariance of two random variables at locations i and j is a decreasing function of an observable distance measure d_{ij} between them. The idea of using a distance measure to characterize spatial dependence is common in the spatial econometrics literature. See, for example, Conley (1999), Kelejian and Prucha (2007), Bester, Conley, Hansen and Vogelsang (2011, BCHV hereafter) and Kim and Sun (2011).

There are several robust covariance estimators with correlated panel data. Arellano (1987) proposes the clustered covariance estimator (CCE) by extending the White standard error (White, 1980) to account for serial correlation. Wooldridge (2003) provides a concise review on the CCE. Driscoll and Kraay (1998, DK hereafter) suggest a different approach that uses a time series HAC estimator (e.g. Newey and West, 1987) applied to cross sectional averages of moment conditions. Gonçalves (2011) examines the properties of this estimator in linear panel models with fixed effects. Another approach considered in this paper is an extension of the spatial HAC estimator applied to time series averages of moment conditions, which we name the DK* estimator. This is symmetric to the original DK estimator. Conley (1999) is among the first to propose the spatial HAC estimator. Kelejian and Prucha (2007) argue that it can be extended to the panel setting with fixed T .

Our estimator includes these existing estimators as special cases, reducing to each of them with certain bandwidth choice. We refer to this as *flexibility*. If the sequence of the bandwidth in the spatial dimension, d_n , increases at a fast enough rate with the cross sectional sample size n , then our estimator with the rectangular kernel is asymptotically equivalent to the DK estimator. Similarly, if the sequence of the bandwidth in the time dimension, d_T , increases fast enough relative to the time series sample size T , then our estimator with the rectangular kernel is asymptotically equivalent to the DK* estimator. On the other hand, if d_n is assumed to approach zero, our estimator reduces to a generalized CCE defined later in the paper.

For distributional approximations, we consider two types of asymptotics: the increasing-smoothing asymptotics and the fixed-smoothing asymptotics. The difference lies in whether the level of smoothing increases or stays fixed as the sample size increases. Let $\ell_{i,n}$ denote the number of individuals whose distance from individual i is less than or equal to d_n and ℓ_n be the average of $\ell_{i,n}$ across i . We also define $\ell_{t,T}$ and ℓ_T in the same way along the time dimension. If $d_n, d_T \rightarrow \infty$ as $(n, T) \rightarrow \infty$ but slowly so that $nT/(\ell_n \ell_T) \rightarrow \infty$, then the level of smoothing increases with the sample size. Under this increasing-smoothing asymptotics, our covariance estimator is consistent and the limiting distribution of the associated Wald statistic is a χ^2 distribution.

The alternative estimators are also consistent under the increasing-smoothing asymptotics, but each estimator has an important limitation in practice. The performance of the CCE heavily depends on spatial correlation. While this estimator is quite efficient in the presence of spatial

independence, even moderate spatial correlation may lead to substantial bias and hence size distortion in statistical testing. Though spatial independence is sometimes assumed for convenience, it may not hold due to, for example, spill-over effects, competition and so on.¹ Collapsing spatial dependence by the cross sectional averaging, the DK estimator is robust to arbitrary forms of spatial dependence when the time series dimension is large. However, when spatial dependence decreases with some distance measure, this estimator is not efficient because it does not down-weight or truncate the covariance between spatially remote units. Similarly, the DK* estimator is not efficient, as it does not employ downweighting or truncation in the time domain.

The proposed estimator improves upon the above estimators by employing a bivariate kernel. It does not require zero spatial correlation for consistency in contrast to the CCE and more efficient than the DK and DK* estimators in general. More specifically, if individuals are located on a 2-dimensional lattice and the Bartlett kernel is used, our estimator is more efficient than the DK estimator if $T = o(n^{3/2})$ and than the DK* estimator if $n = o(T^4)$. For second-order kernels, the conditions become much weaker, i.e. $T = o(n^{5/2})$ and $n = o(T^6)$, respectively.

If we embed the bandwidth parameters d_n and d_T in a sequence such that $nT/(\ell_n \ell_T)$ holds fixed as n and T increase, then the level of smoothing is fixed with the sample size. Under this fixed-smoothing asymptotics, the Wald statistic is asymptotically equivalent to a nonstandard but pivotal distribution. The fixed-smoothing asymptotic approximation is first suggested by Kiefer, Vogelsang and Bunzel (2000) and Kiefer and Vogelsang (2002a, 2002b, 2005) in the time series context. This is usually referred to as the ‘fixed- b ’ asymptotics where b denotes the ratio of the bandwidth parameter d_T to the sample size T . They show by simulation that the fixed- b asymptotic approximation is more accurate than the χ^2 approximation. Jansson (2004), Sun, Phillips and Jin (2008), and Sun and Phillips (2009) provide theoretical explanations in different time series settings.

We adopt the fixed-smoothing asymptotics in the panel setting with our covariance estimator. Based on a CLT, we first show that, under the fixed smoothing asymptotics, the Wald statistic is asymptotically equivalent in distribution to a quadratic form of a standard normal vector with an independent and random weighting matrix. Using a CLT is an important departure from the previous literature in which the fixed-smoothing asymptotics relies on an FCLT. The CLT holds under mild conditions, so our asymptotic results are applicable to a wide range of panel data processes. This is one of the theoretical contributions of our paper.

Using asymptotic expansions, we show that the deviation of the asymptotically equivalent distribution from the χ^2 distribution depends on the smoothing parameters, the kernel function, and the number of restrictions being tested. We can account for the estimation uncertainty of the parameter estimation and the randomness of the covariance estimator under the fixed-smoothing asymptotics. To conduct hypothesis testing and construct confidence intervals, we can simulate the asymptotically equivalent distribution. The asymptotically equivalent distribution is nonstandard but a function of nT *i.i.d.* standard normal vectors, so it is easy to simulate. We also extend Sun (2010) to establish the validity of an F -approximation to the distribution. We show that the asymptotically equivalent distribution of the scaled Wald statistic with some correction factor becomes approximately F distributed. This F -approximation greatly facilitates

¹Recently, Bester, Conley and Hansen (2011) present consistency results for the CCE with spatially dependent data by constructing clusters to be asymptotically independent. In this paper, we consider a rather traditional panel CCE for which the cluster is defined based on each individual so that the asymptotic independence condition is not valid. Cameron, Gelbach, and Miller (2011) address this problem by clustering on the time and spatial dimensions simultaneously. While this allows for both the serial and spatial correlations, observations on different individuals in different time are assumed to be uncorrelated.

the testing procedure and yields accurate critical values when the bandwidths are small.

Several testing methods using the fixed-smoothing asymptotics are recently proposed in the spatial or panel setting. BCHV extend the fixed- b asymptotics to the spatial context where dependence is indexed in more than one dimension, and propose an *i.i.d.* bootstrap method to obtain the critical values. Vogelsang (2012) develops a fixed- b asymptotic theory for statistics based on the generalized CCE and the DK estimator. Besides the kernel methods, Hansen (2007) and Bester, Conley and Hansen (2011) apply the fixed-smoothing asymptotics to the testing procedure with the CCE. They assume the number of clusters to be fixed and the number of observations per cluster to increase with the sample size. Ibragimov and Müller (2010) consider the fixed-smoothing asymptotics for the Fama and MacBeth (1973) type procedure by fixing the number of groups. Sun and Kim (2013) consider a testing procedure using a series-type covariance estimator in the spatial setting. They show that, when the number of basis functions is held fixed, the scaled Wald statistic with the series covariance estimator is asymptotically equivalent in distribution to an F distribution. Our F -approximation is motivated from the series method of Sun and Kim (2012). While for the other ‘non-kernel’ methods critical values are readily available from the standard t or F distribution, critical values for the kernel methods by BCHV and Vogelsang (2012) have to be simulated. From this point of view, this paper fills the gap in the literature, providing an F -approximation for the kernel method in the panel setting.

In this paper, we select the bandwidth parameters to minimize an upper bound of an approximate mean square error (called AMSE*) of the covariance estimator. The AMSE* criterion has a minimax flavor. It is simple to implement and makes the bias and variance tradeoff transparent. It is interesting to note that the level of persistence in each dimension affects both d_T^* and d_n^* , the optimal bandwidth parameters in the time and spatial dimensions respectively, but in opposite directions. We suggest a parametric plug-in procedure for practical implementation using the spatiotemporal models in Anselin (2001).

Our bandwidth selection procedure does not apply directly to the rectangular kernel estimator and, more broadly, flat-top kernel estimators. However, it is interesting to consider flat-top kernel estimators because they are higher-order accurate (Politis, 2011). This is particularly important in our setting because the rectangular-kernel-based covariance estimator is more flexible in that it can approach each of the existing estimators with appropriate bandwidth choice. We modify our bandwidth selection procedure to be applicable to flat-top kernels which include the rectangular kernel as a special case. The rectangular kernel, combined with our modified bandwidth selection procedure, delivers a covariance estimator with better asymptotic properties than the covariance estimators based on second-order kernels.

The flexibility of our covariance estimator and the data-driven bandwidth selection procedure make our estimator adaptive to the dependence structure in the data. That is, in large samples, our estimator reduces to the estimator that is designed to cope with a particular dependence structure. This *adaptiveness* is the salient feature of our method. As it practically automates the selection of covariance estimators, our estimation procedure can be safely used in the presence of very general forms of spatiotemporal dependence. This is confirmed by our Monte Carlo simulation study.

The remainder of the paper is as follows. Section 2 introduces the panel model, the covariance estimator and hypothesis testing we consider. In Section 3, we examine the properties of our estimator and the associated test statistic under the increasing-smoothing asymptotics. Section 4 develops an optimal bandwidth selection procedure. Section 5 examines the properties of the existing estimators. The flexibility and adaptiveness of our estimator are illustrated in Section 6.

In Section 7, we develop the asymptotic theory for our covariance estimator and the associated test statistic under the fixed-smoothing asymptotics. We also prove the validity of an F -approximation to the Wald statistic. Section 8 reports simulation evidence. The last section concludes. Proofs are given in the appendix or in a supplementary appendix at the end of the paper. All limits for $(n, T) \rightarrow \infty$ in the paper are taken as joint limits.

2 Panel model, covariance estimator and hypothesis testing

In this paper, we consider a linear panel regression model with fixed effects given by

$$Y_{it} = X'_{it}\beta_0 + \alpha_i + f_t + u_{it} \quad (1)$$

where α_i and f_t denote individual and time effects respectively. Either α_i or f_t has a nonzero mean so there is no constant regressor in X_{it} . We allow the p covariates in X_{it} to be correlated with α_i , $\{f_t\}$ and $\{u_{it}\}$. We assume that there are $d_z \geq p$ instrumental variables Z_{it} which satisfy the conditions in Assumption 1 below.

When Z_{it} is correlated with α_i and f_t , we may use the fixed effects 2SLS estimator. Let $\bar{Z}_i = T^{-1} \sum_{t=1}^T Z_{it}$, $\bar{Z}_t = n^{-1} \sum_{i=1}^n Z_{it}$ and $\bar{Z} = (nT)^{-1} \sum_{i=1}^n \sum_{t=1}^T Z_{it}$. Define $\tilde{Z}_{it} = Z_{it} - \bar{Z}_i - \bar{Z}_t + \bar{Z}$ and apply the same definition to X_{it} and other variables. Then, the fixed effects 2SLS estimator $\hat{\beta}$ of β is given by

$$\begin{aligned} \hat{\beta} = & \left(\left(\sum_{i=1}^n \sum_{t=1}^T \tilde{X}_{it} \tilde{Z}'_{it} \right) \left(\sum_{i=1}^n \sum_{t=1}^T \tilde{Z}_{it} \tilde{Z}'_{it} \right)^{-1} \sum_{i=1}^n \sum_{t=1}^T \tilde{Z}_{it} \tilde{X}'_{it} \right)^{-1} \\ & \times \left(\sum_{i=1}^n \sum_{t=1}^T \tilde{X}_{it} \tilde{Z}'_{it} \right) \left(\sum_{i=1}^n \sum_{t=1}^T \tilde{Z}_{it} \tilde{Z}'_{it} \right)^{-1} \sum_{i=1}^n \sum_{t=1}^T \tilde{Z}_{it} Y_{it}. \end{aligned} \quad (2)$$

When the underlying probability limits are well defined, we have, under some mild conditions:

$$\sqrt{nT} (\hat{\beta} - \beta_0) = \left(\frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \tilde{X}_{it} \tilde{Z}'_{it} Q_{ZZ}^{-1} Q_{ZX} \right)^{-1} \frac{1}{\sqrt{nT}} \sum_{i=1}^n \sum_{t=1}^T Q_{XZ} Q_{ZZ}^{-1} \tilde{Z}_{it} u_{it} + o_p(1),$$

where $Q_{ZX} = \text{plim}_{(n,T) \rightarrow \infty} (nT)^{-1} \sum_{i,t} \tilde{Z}_{it} \tilde{X}'_{it}$, $Q_{ZZ} = \text{plim}_{(n,T) \rightarrow \infty} (nT)^{-1} \sum_{i,t} \tilde{Z}_{it} \tilde{Z}'_{it}$ and $Q_{XZ} = Q'_{ZX}$. So it does not matter in the first order asymptotics whether we use Z_{it} as the instruments or their linear combinations $Q_{XZ} Q_{ZZ}^{-1} Z_{it}$ as the instruments. While the number of the original instruments Z_{it} may be larger than the number of endogenous covariates X_{it} , the numbers of the transformed instruments $Q_{XZ} Q_{ZZ}^{-1} Z_{it}$ and X_{it} are exactly the same by construction. For the sake of notational simplicity and clarity, we will assume from now on that the number of instruments is the same as the number of endogenous covariates, i.e. $p = d_z$. For the asymptotic properties we are interested in here, we do so without loss of generality.² With this assumption, we have

$$\sqrt{nT} (\hat{\beta} - \beta_0) = Q^{-1} \frac{1}{\sqrt{nT}} \sum_{i=1}^n \sum_{t=1}^T \tilde{Z}_{it} u_{it} + o_p(1)$$

where we have written $Q := Q_{ZX}$.

²We do not consider the weak IV or many (weak) IV asymptotics here.

Define $\check{Z}_{it} = Z_{it} - \mu_{zi} - \mu_{zt} + \mu_z$ where μ_{zi} , μ_{zt} and μ_z satisfy $\text{plim}_{T \rightarrow \infty} T^{-1} \sum_t Z_{it} - \mu_{zi} = 0$, $\text{plim}_{n \rightarrow \infty} n^{-1} \sum_i Z_{it} - \mu_{zt} = 0$, and $\text{plim}_{(n,T) \rightarrow \infty} (nT)^{-1} \sum_{i,t} Z_{it} - \mu_z = 0$. While \check{Z}_{it} is the sample demeaned version of Z_{it} , \check{Z}_{it} can be regarded as its population demeaned analogue. If $Z_{it} = Z_{it}^o + \alpha_{zi} + f_{zt}$ for some stationary and weakly dependent spatiotemporal process Z_{it}^o , spatial process α_{zi} and time series f_{zt} such that $\text{plim}_{T \rightarrow \infty} T^{-1} \sum_t Z_{it}^o = \text{plim}_{n \rightarrow \infty} n^{-1} \sum_i Z_{it}^o = \text{plim}_{(n,T) \rightarrow \infty} (nT)^{-1} \sum_{i,t} Z_{it}^o = E Z_{it}^o$ and $\text{plim}_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \alpha_{zi}$ and $\text{plim}_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T f_{zt}$ exist, then $\check{Z}_{it} = Z_{it}^o - E Z_{it}^o$, a weakly dependent spatiotemporal process with mean zero.

Let

$$J_{nT} = \text{var} \left((nT)^{-1/2} \sum_{i=1}^n \sum_{t=1}^T \check{Z}_{it} u_{it} \right).$$

We make some assumptions on the instruments.

Assumption 1 (i) *There is a nonsingular matrix Q such that $Q = \text{plim}_{(n,T) \rightarrow \infty} (nT)^{-1} \sum_{i,t} \check{Z}_{it} \check{X}_{it}'$.*

(ii) $\frac{1}{\sqrt{nT}} \sum_{i=1}^n \sum_{t=1}^T \check{Z}_{it} u_{it} = \frac{1}{\sqrt{nT}} \sum_{i=1}^n \sum_{t=1}^T \check{Z}_{it} u_{it} + o_p(1)$.

(iii) $J_{nT}^{-1/2} \frac{1}{\sqrt{nT}} \sum_{i=1}^n \sum_{t=1}^T \check{Z}_{it} u_{it} \rightarrow^d N(0, I_p)$.

Assumption 1(i) and (iii) hold by LLN and CLT respectively. Assumption 1(iii) implicitly assumes that for each individual there is no contemporaneous correlation between Z_{it} and u_{it} . This is a minimal condition that valid instruments have to satisfy.

Assumption 1(ii) requires that finite sample demeaning does not affect the first order asymptotic distribution. Since

$$\check{Z}_{it} = \check{Z}_{it} - (\bar{Z}_i - \mu_{zi}) - (\bar{Z}_t - \mu_{zt}) + (\bar{Z} - \mu_z),$$

Assumption 1(ii) is equivalent to

$$\frac{1}{\sqrt{nT}} \sum_{i=1}^n \sum_{t=1}^T (\bar{Z}_i - \mu_{zi}) u_{it} = o_p(1), \quad (3)$$

$$\frac{1}{\sqrt{nT}} \sum_{i=1}^n \sum_{t=1}^T (\bar{Z}_t - \mu_{zt}) u_{it} = o_p(1), \quad (4)$$

$$\frac{1}{\sqrt{nT}} \sum_{i=1}^n \sum_{t=1}^T (\bar{Z} - \mu_z) u_{it} = o_p(1). \quad (5)$$

A sufficient condition for the above to hold is that each left hand side (lhs) has mean and variance approaching zero. We focus on the mean here. For the lhs of (3), we have

$$\begin{aligned} E \frac{1}{\sqrt{nT}} \sum_{i=1}^n \sum_{t=1}^T (\bar{Z}_i - \mu_{zi}) u_{it} &= \frac{1}{\sqrt{nT}} \sum_{i=1}^n \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T E (Z_{it} - \mu_{zi}) u_{is} \\ &= \sqrt{\frac{n}{T}} \frac{1}{n} \sum_{i=1}^n \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T E (Z_{it} - \mu_{zi}) u_{is} := \sqrt{\frac{n}{T}} \frac{1}{n} \sum_{i=1}^n \Pi_i, \end{aligned}$$

where $n^{-1} \sum_{i=1}^n \Pi_i$ is the cross sectional average of the time series long run covariance³ between $Z_{it} - \mu_{zi}$ and u_{it} . Sufficient conditions for the mean to diminish are (i) $n/T \rightarrow 0$ and

³Strictly speaking, for each individual i , the long run covariance between $Z_{it} - \mu_{zi}$ and u_{it} is $\lim_{T \rightarrow \infty} \Pi_i$ so Π_i should be viewed as a finite sample version of the long run covariance.

$\|n^{-1} \sum_{i=1}^n \Pi_i\| \leq C_\Pi < \infty$ some constant C_Π and all n or (ii) $n/T \leq C < \infty$ for some constant C and all (n, T) and $\|n^{-1} \sum_{i=1}^n \Pi_i\| \rightarrow 0$. For the lhs of (4), we have

$$E \frac{1}{\sqrt{nT}} \sum_{i=1}^n \sum_{t=1}^T (\bar{Z}_t - \mu_{zt}) u_{it} = \sqrt{\frac{T}{n}} \frac{1}{T} \sum_{t=1}^T \Pi_t, \quad (6)$$

where $\Pi_t = n^{-1} \sum_{i=1}^n \sum_{j=1}^n E[(Z_{it} - \mu_{zt}) u_{jt}]$ can be regarded as the cross sectional ‘long run’ covariance between $Z_{it} - \mu_{zt}$ and u_{it} . For the mean in (6) to diminish, we require (i) $T/n \rightarrow 0$ and $T^{-1} \sum_{t=1}^T \Pi_t \leq C_\Pi < \infty$ for some finite constant C_Π and all T or (ii) $T/n \leq C < \infty$ for some constant C and all (n, T) and $T^{-1} \sum_{t=1}^T \Pi_t \rightarrow 0$. Finally, for the lhs of (5), we have

$$E \frac{1}{\sqrt{nT}} \sum_{i=1}^n \sum_{t=1}^T (\bar{Z} - \mu_z) u_{it} = \frac{1}{\sqrt{nT}} \text{var} \left(\frac{1}{\sqrt{nT}} \sum_{i=1}^n \sum_{t=1}^T Z_{it} u_{it} \right).$$

This indeed approaches zero under some moment and mixing conditions.

To sum up, a set of sufficient conditions for Assumption 1(ii) is:

- (a) the cross sectional average of the time series long run covariance between $Z_{it} - \mu_{zi}$ and u_{it} vanishes;
- (b) the time series average of the cross sectional ‘long run’ covariance between $Z_{it} - \mu_{zt}$ and u_{it} vanishes;
- (c) there are enough moment and mixing conditions.

When Z_{it} is strongly exogenous so that $\text{cov}(Z_{it}, u_{js}) = 0$ for all i, t, j, s , conditions (a) and (b) are satisfied automatically. When Z_{it} is sequentially exogenous in the time dimension and strictly exogenous in the cross sectional dimension so that $\text{cov}(Z_{it}, u_{js}) = 0$ for all i, j and $t \leq s$, condition (b) holds but condition (a) does not. In this case, if $n/T \rightarrow 0$, then Assumption 1(ii) still holds. So our assumption can accommodate some dynamic models.

When Assumption 1(ii) does not hold, there will be a first order bias in $\hat{\beta}$, arising from the incidental parameters problem first considered by Neyman and Scott (1948). That is, $\sqrt{nT}(\hat{\beta} - \beta_0)$ will not be centered at zero even in large samples. In this case, a bias correction procedure will be needed for confidence interval construction and hypothesis testing. In the panel setting with large n and T , there are bias correction procedures that do not change the asymptotic variance of $\hat{\beta}$ under some conditions. See for example, Arellano and Hahn (2006) and references therein. For this reason, we will not pursue bias correction here and focus only on variance estimation.

Under Assumption 1, the asymptotic distribution of $\hat{\beta}$ is

$$(Q^{-1} J_{nT} Q'^{-1})^{-1/2} \sqrt{nT} (\hat{\beta} - \beta_0) \xrightarrow{d} N(0, I_p) \text{ as } (n, T) \rightarrow \infty.$$

To make inference on β_0 , we have to estimate unknown quantities in the asymptotic variance of $\hat{\beta}$. Since Q can be consistently estimated by its sample analog $\hat{Q}_{nT} := (nT)^{-1} \sum_{i,t} \tilde{Z}_{it} \tilde{X}'_{it}$, our central interest is on J_{nT} . Let $V_{(i,t)} = \tilde{Z}_{it} u_{it}$, then J_{nT} can be rewritten as

$$J_{nT} = \frac{1}{nT} \sum_{i,j=1}^n \sum_{t,s=1}^T E[V_{(i,t)} V'_{(j,s)}].$$

We propose a bivariate kernel covariance estimator given by⁴

$$\hat{J}_{nT} = \frac{1}{nT} \sum_{i,j=1}^n \sum_{t,s=1}^T K_1 \left(\frac{d_{ij}}{d_n} \right) K_2 \left(\frac{d_{ts}}{d_T} \right) \hat{V}_{(i,t)} \hat{V}_{(j,s)}', \quad (7)$$

where $\hat{V}_{(i,t)} = \tilde{Z}_{it}(\tilde{Y}_{it} - \tilde{X}_{it}'\hat{\beta})$, $K_1(\cdot)$ and $K_2(\cdot)$ are real-valued kernel functions, and d_{ij}, d_{ts} and d_n, d_T denote the corresponding distance measures and bandwidth parameters. For simplicity, we have used the product kernel $K_1(\cdot)K_2(\cdot)$ in the above covariance estimator. Whereas it is natural to define $d_{ts} = |t - s|$, what is used to measure d_{ij} differs with applications. Geographic distance is one of the most common measures, but other measures can also be considered, e.g. transportation cost (Conley and Ligon, 2000) and similarity of input and output structure (Chen and Conley, 2001; and Conley and Dupor, 2003).

Consider the null hypothesis $H_0 : \mathcal{R}\beta = r_0$ and alternative hypothesis $H_1 : \mathcal{R}\beta \neq r_0$ where \mathcal{R} is a $g \times p$ matrix and r_0 is a g -vector. For hypothesis testing, we use the Wald statistic

$$\mathcal{W}_{nT} = \sqrt{nT} \left(\mathcal{R}\hat{\beta} - r_0 \right)' \left(\mathcal{R}\hat{Q}_{nT}^{-1} \hat{J}_{nT} \hat{Q}_{nT}^{-1} \mathcal{R}' \right)^{-1} \sqrt{nT} \left(\mathcal{R}\hat{\beta} - r_0 \right)$$

and its F -test version $F_{nT} = \mathcal{W}_{nT}/g$. With some obvious modification, a t test can also be performed. Our results remain valid for nonlinear restrictions after linearization.

3 Increasing-smoothing asymptotics

3.1 Basic setting

There are nTp variables in $\{V_{(i,t)}, i = 1, \dots, n \text{ and } t = 1, \dots, T\}$. Inspired by KP, we consider the following linear array process to represent $V_{(i,t)}$. We assume that each element of $V_{(i,t)}$ responds linearly to nTp common innovations $\{\tilde{\varepsilon}_{(i,t)}^{(c)}, i = 1, \dots, n, t = 1, \dots, T, c = 1, \dots, p\}$:

$$V_{(i,t)} = \tilde{R}_{(i,t)} \tilde{\varepsilon}, \quad (8)$$

where

$$\tilde{R}_{(i,t)} = \begin{bmatrix} \left(\tilde{r}_{(it,1,1)}^{(1)}, \tilde{r}_{(it,2,1)}^{(1)}, \dots, \tilde{r}_{(it,n,T)}^{(1)} \right) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \left(\tilde{r}_{(it,1,1)}^{(p)}, \tilde{r}_{(it,2,1)}^{(p)}, \dots, \tilde{r}_{(it,n,T)}^{(p)} \right) \end{bmatrix}$$

is a $p \times nTp$ block diagonal matrix with unknown elements, and $\tilde{\varepsilon} = (\tilde{\varepsilon}^{(1)'}, \dots, \tilde{\varepsilon}^{(p)'})'$ in which $\tilde{\varepsilon}^{(c)} = (\tilde{\varepsilon}_{(1,1)}^{(c)}, \dots, \tilde{\varepsilon}_{(n,1)}^{(c)}, \tilde{\varepsilon}_{(1,2)}^{(c)}, \dots, \tilde{\varepsilon}_{(n,T)}^{(c)})'$. The block diagonal elements, $\tilde{r}_{(it,j,s)}^{(c)}$, are implicitly allowed to depend on n and T . We also define $\tilde{R}_{(i,t)}^{(c)} = (\tilde{r}_{(it,1,1)}^{(c)}, \tilde{r}_{(it,2,1)}^{(c)}, \dots, \tilde{r}_{(it,n,T)}^{(c)})$. As in Kim and Sun (2011), we assume

$$\text{var}(\tilde{\varepsilon}^{(c)}) = \sigma_{cc} I_{nT}, \text{cov}(\tilde{\varepsilon}^{(c)}, \tilde{\varepsilon}^{(d)}) = \sigma_{cd} I_{nT}$$

⁴For notational economy, here we have abused the notation by using the same d to denote the distances along the time and spatial dimensions. This should not cause any confusion.

and

$$\text{var}(\tilde{\varepsilon}) = \Sigma \otimes I_{nT} \text{ with } \Sigma = (\sigma_{cd}),$$

where $c, d = 1, \dots, p$ and \otimes denotes the Kronecker product. This type of linear array processes allows for nonstationarity and unconditional heteroskedasticity of $V_{(i,t)}$ and includes many spatiotemporal parametric models such as spatial dynamic models (Anselin, 2001) as special cases. It also treats the temporal and spatial dependence in a symmetric way.

Let $R_{(i,t)} := \tilde{R}_{(i,t)} (\Sigma^{1/2} \otimes I_{nT})$ and $\varepsilon := (\varepsilon_1, \dots, \varepsilon_l, \dots, \varepsilon_{nTp})' = (\Sigma^{-1/2} \otimes I_{nT}) \tilde{\varepsilon}$. Then,

$$V_{(i,t)} = R_{(i,t)} \varepsilon \text{ and } \text{var}(\varepsilon) = I_{nTp}. \quad (9)$$

The matrix $R_{(i,t)}$ can be written more explicitly as

$$\begin{aligned} R_{(i,t)} &:= \begin{bmatrix} \left(\begin{array}{ccc} r_{(i,t),1}^{(1)} & \cdots & r_{(i,t),nTp}^{(1)} \end{array} \right) \\ \vdots \\ \left(\begin{array}{ccc} r_{(i,t),1}^{(p)} & \cdots & r_{(i,t),nTp}^{(p)} \end{array} \right) \end{bmatrix} \\ &= \begin{bmatrix} \sigma^{11} \left(\begin{array}{ccc} \tilde{r}_{(it,1,1)}^{(1)} & \cdots & \tilde{r}_{(it,n,T)}^{(1)} \end{array} \right) & \cdots & \sigma^{1p} \left(\begin{array}{ccc} \tilde{r}_{(it,1,1)}^{(1)} & \cdots & \tilde{r}_{(it,n,T)}^{(1)} \end{array} \right) \\ \vdots & \ddots & \vdots \\ \sigma^{p1} \left(\begin{array}{ccc} \tilde{r}_{(it,1,1)}^{(p)} & \cdots & \tilde{r}_{(it,n,T)}^{(p)} \end{array} \right) & \cdots & \sigma^{pp} \left(\begin{array}{ccc} \tilde{r}_{(it,1,1)}^{(p)} & \cdots & \tilde{r}_{(it,n,T)}^{(p)} \end{array} \right) \end{bmatrix} \end{aligned}$$

where σ^{cd} denotes the (c, d) -th element of $\Sigma^{1/2}$. We also define the c -th row vector of $R_{(i,t)}$ as $R_{(i,t)}^{(c)} = (r_{(i,t),1}^{(c)}, \dots, r_{(i,t),nTp}^{(c)})$. We make the following assumption on ε_l .

Assumption I1 For all $l = 1, \dots, nTp$, $\varepsilon_l \stackrel{i.i.d.}{\sim} (0, 1)$ with $E[\varepsilon_l^4] \leq c_E$ for some constant $c_E < \infty$.

For simplicity, we assume that ε_l is independent of ε_k for $l \neq k$. We can relax the independence assumption to zero correlation which holds by construction but with more tedious calculations.

Under Assumption I1, the covariance matrix of $V_{(i,t)}$ and $V_{(j,s)}$ is given by

$$\Gamma_{(it,js)} := \left(\gamma_{(it,js)}^{(cd)} \right) = E \left[V_{(i,t)} V_{(j,s)}' \right] = R_{(i,t)} R_{(j,s)}', \quad (10)$$

where the (c, d) -th element of $\Gamma_{(it,js)}$ is denoted by $\gamma_{(it,js)}^{(cd)}$. Accordingly, J_{nT} can be rewritten as

$$J_{nT} = \frac{1}{nT} \sum_{i,j=1}^n \sum_{t,s=1}^T R_{(i,t)} R_{(j,s)}',$$

and the (c, d) -th element of J_{nT} is

$$J_{nT}(c, d) = \frac{1}{nT} \sum_{i,j=1}^n \sum_{t,s=1}^T R_{(i,t)}^{(c)} \left(R_{(j,s)}^{(d)} \right)'.$$

Assumption I2 For all $l = 1, \dots, nTp$, $c = 1, \dots, p$, and all (n, T) , $\sum_{i=1}^n \sum_{t=1}^T \left| r_{(i,t),l}^{(c)} \right| < c_R$ for some constant c_R , $0 < c_R < \infty$.

Assumption I3 *There exist finite positive constants q_S, q_T, c_S and c_T such that*

$$(i) \frac{1}{nT} \sum_{i,j=1}^n \sum_{t,s=1}^T \|\Gamma_{(it,j,s)}\| d_{ij}^{q_S} < c_S \text{ and } (ii) \frac{1}{nT} \sum_{i,j=1}^n \sum_{t,s=1}^T \|\Gamma_{(it,j,s)}\| d_{ts}^{q_T} < c_T,$$

for all (n, T) , where $\|A\|$ denotes the Euclidean norm of matrix A .

Assumptions I2 and I3 impose the conditions on the persistence of the process. If for all c and d , $|\sigma^{cd}| \leq c_\sigma$ for a constant $c_\sigma > 0$, then Assumption I2 holds if $\sum_{i=1}^n \sum_{t=1}^T |\tilde{r}_{(it,j,s)}^{(c)}| < c_R/c_\sigma$ for all n and T . Since $|\tilde{r}_{(it,j,s)}^{(c)}|$ can be regarded as the (absolute) change of $V_{(i,t)}^{(c)}$ in response to one unit change in one element of $\tilde{\varepsilon}^{(c)}$, the summability condition requires that the aggregate response to an innovation be finite. Assumption I3 implies that $\Gamma_{(it,j,s)}$ decays to zero fast enough as d_{ij} and d_{ts} increase so that the two summability conditions hold. These conditions hold if

$$\limsup_{(n,T) \rightarrow \infty} \frac{1}{nT} \sum_{i,j=1}^n \sum_{t,s=1}^T \|\tilde{R}_{(i,t)}^{(c)} \left(\tilde{R}_{(j,s)}^{(d)} \right)'\| d_{ij}^{q_S} < \infty, \quad (11)$$

$$\limsup_{(n,T) \rightarrow \infty} \frac{1}{nT} \sum_{i,j=1}^n \sum_{t,s=1}^T \|\tilde{R}_{(i,t)}^{(c)} \left(\tilde{R}_{(j,s)}^{(d)} \right)'\| d_{ts}^{q_T} < \infty \quad (12)$$

for all c and d . (11) and (12) imply that as d_{ij} or d_{ts} increases, the corresponding two row vectors $\tilde{R}_{(i,t)}^{(c)}$ and $\tilde{R}_{(j,s)}^{(d)}$ become nearly orthogonal. As the row vector represents the aggregate response of a unit to all the innovations, this assumption implies that the responses of two units become independent as they become spatially or temporally distant. Assumption I3 enables us to truncate the sum of $\Gamma_{(it,j,s)}$ and downweigh the summand without incurring much bias.

As Assumption I3 implies, the key property of d_{ij} is to characterize the decaying pattern of the spatial dependence. In addition, we assume that d_{ij} satisfies the properties of a distance measure in a metric space: (i) $d_{ij} \geq 0$, (ii) $d_{ii} = 0$, (iii) $d_{ij} = d_{ji}$, and (iv) $d_{ij} \leq d_{ik} + d_{kj}$. In practice, nonetheless, the symmetry condition (iii) may not hold for some candidates of economic distance. Conley and Ligon (2000), for example, notice that transportation costs among countries violate this condition if tariff barriers are asymmetric. In such a case adjustment should be made.⁵ This adjustment does not affect the asymptotic properties of our estimator from the perspective of the measurement error problem as we now explain.

Distance measures observable to empirical researchers usually contain measurement errors, and the results in this paper can be generalized to the case when d_{ij} is error contaminated. Following Kim and Sun (2011), we can show that our asymptotic results are still valid under the following conditions: (i) the measurement error is independent of ε_l for all l ; (ii) it is of order $o(d_n)$ as d_n increases; and (iii) the summability condition in Assumption I3(i) holds with the error-contaminated distance measure. In this paper, however, we do not consider measurement errors for simplicity.

Let

$$\ell_{i,n} = \sum_{j=1}^n 1\{d_{ij} \leq d_n\} \text{ and } \ell_n = n^{-1} \sum_{i=1}^n \ell_{i,n}.$$

⁵In Conley and Ligon (2000), an asymmetric transportation cost is replaced by the minimum cost between two countries.

$\ell_{i,n}$ is the number of pseudo-neighbors that unit i has and ℓ_n is the average number of pseudo-neighbors. Here we use the terminology “pseudo-neighbor” in order to differentiate it from the common usage of “neighbor” in spatial modeling. We maintain the following assumption on the number of pseudo-neighbors.

Assumption I4 For all $i = 1, \dots, n$, $\ell_{i,n} \leq c\ell_n$ for some constant c .

Assumption I4 allows the units to be irregularly located but rules out the case that they are concentrated only in some limited areas. To be symmetric, we also define

$$\ell_{t,T} = \sum_{s=1}^T 1\{|t-s| \leq d_T\} \text{ and } \ell_T = T^{-1} \sum_{t=1}^T \ell_{t,T} = 2d_T + 1 - \frac{d_T(d_T+1)}{T},$$

where $-d_T(d_T+1)/T$ is an adjustment coming from the points near the boundary.

In order to obtain the properties of the estimator in Theorem 1 below, it is important to control for the boundary effects. It is especially critical in the panel setting, because it faces a larger boundary than the time series and spatial settings. The effects of the units near the boundary should become negligible as the sample size increases so that the asymptotic properties depend only on the behavior of the units in the interior. We define

$$E_n := \{i : \ell_{i,n} = \ell_n + o(\ell_n)\}, \quad n_1 = \sum_{i=1}^n 1\{i \in E_n\}, \quad n_2 = n - n_1$$

$$E_T := \{t : \ell_{t,T} = \ell_T + o(\ell_T)\}, \quad T_1 = \sum_{t=1}^T 1\{t \in E_T\} \text{ and } T_2 = T - T_1.$$

E_n and E_T represent the nonboundary sets in the spatial and time dimensions. n_1 and T_1 denote the sizes of E_n and E_T and n_2 and T_2 denote the sizes of the boundary sets. These definitions imply that the size of a boundary set depends on choice of the bandwidth parameters. We can mitigate the boundary effects by raising d_n and d_T slowly as n and T increase to make the interior large enough. Provided that n_2/n and T_2/T are $o(1)$, the boundary effects are asymptotically negligible. When units are regularly spaced on a lattice in \mathbb{R}^2 , $n_2/n = o(1)$ if $\ell_n/n = o(1)$. $T_2/T = o(1)$ holds if $\ell_T/T = o(1)$.

3.2 Increasing-smoothing asymptotics

In this subsection, we investigate the asymptotic properties of \hat{J}_{nT} and the limiting distribution of the Wald statistic \mathcal{W}_{nT} under the increasing-smoothing asymptotics.

Following the standard practice, we could define the (normalized) MSE of \hat{J}_{nT} as

$$MSE\left(\frac{nT}{\ell_n \ell_T}, \hat{J}_{nT}, S\right) = \frac{nT}{\ell_n \ell_T} E\left[\text{vec}(\hat{J}_{nT} - J_{nT})' S \text{vec}(\hat{J}_{nT} - J_{nT})\right],$$

where S is some $p^2 \times p^2$ positive definite weighting matrix⁶ and $\text{vec}(\cdot)$ is the column by column vectorization function. However, the mean and variance of \hat{J}_{nT} may not exist. For example, when the model is exactly identified, $\hat{\beta}$ has no moment (Mariano, 1972). A direct implication

⁶The weighting matrix may depend on (n, T) in which case we assume that $S_{nT} \rightarrow S$ as $(n, T) \rightarrow \infty$ and our asymptotic results remain valid.

is that the above MSE is not well defined for an exactly identified model or an over-identified model with only one over-identifying condition.

To overcome this technical difficulty, we introduce the pseudo-estimator:

$$\tilde{J}_{nT} = \frac{1}{nT} \sum_{i,j=1}^n \sum_{t,s=1}^T K_1\left(\frac{d_{ij}}{d_n}\right) K_2\left(\frac{d_{ts}}{d_T}\right) V_{(i,t)} V'_{(j,s)},$$

which is identical to \hat{J}_{nT} but with sample statistics replaced by their population analogue. If $\hat{J}_{nT} - \tilde{J}_{nT} = o_p(\sqrt{\ell_n \ell_T / (nT)})$, then we can use $MSE(nT / (\ell_n \ell_T), \tilde{J}_{nT}, S)$ as an approximate MSE for \hat{J}_{nT} . This is a Nagar type of approximation (Nagar, 1959). The approximation can be justified to some extent using the truncation argument of Andrews (1991). For any $h > 0$, let

$$MSE_h\left(\frac{nT}{\ell_n \ell_T}, \hat{J}_{nT}, S\right) = \frac{nT}{\ell_n \ell_T} E\left[\min\left(\text{vec}(\hat{J}_{nT} - J_{nT})' S \text{vec}(\hat{J}_{nT} - J_{nT}), h\right)\right]$$

be the truncated MSE, which exists for any h by construction. Then under some conditions, it can be shown that

$$\lim_{h \rightarrow \infty} \lim_{(n,T) \rightarrow \infty} MSE_h\left(\frac{nT}{\ell_n \ell_T}, \hat{J}_{nT}, S\right) = \lim_{h \rightarrow \infty} \lim_{(n,T) \rightarrow \infty} MSE_h\left(\frac{nT}{\ell_n \ell_T}, \tilde{J}_{nT}, S\right).$$

Andrews (1991) and Kim and Sun (2011) provide the conditions for the above to hold in the time series and spatial settings, respectively. In this paper, we make the following high level assumption, whose sufficient conditions are given in the supplementary appendix.

Assumption I5 $\hat{J}_{nT} - \tilde{J}_{nT} = o_p(\sqrt{\ell_n \ell_T / (nT)})$.

Under this assumption, we employ $MSE(nT / (\ell_n \ell_T), \tilde{J}_{nT}, S)$ directly as the Nagar-type approximate MSE (AMSE) of \hat{J}_{nT} . We define

$$AMSE\left(\frac{nT}{\ell_n \ell_T}, \hat{J}_{nT}, S\right) := MSE\left(\frac{nT}{\ell_n \ell_T}, \tilde{J}_{nT}, S\right).$$

To compute $MSE(nT / (\ell_n \ell_T), \tilde{J}_{nT}, S)$, we introduce the assumption below.

Assumption I6 (i) $\ell_n/n = o(1)$, $\ell_T/T = o(1)$, $d_n \rightarrow \infty$ and $d_T \rightarrow \infty$ as $(n, T) \rightarrow \infty$, (ii) for $i \in E_n$ and $t \in E_T$,

$$\lim_{(n,T) \rightarrow \infty} \text{var}\left(\frac{1}{\sqrt{\ell_{i,n} \ell_{t,T}}} \sum_{j: d_{ij} \leq d_n} \sum_{s: d_{ts} \leq d_T} V_{(j,s)}\right) = \lim_{(n,T) \rightarrow \infty} J_{nT} := J.$$

Assumption I6 states that the covariance matrix defined locally for each nonboundary unit converges to the same limiting value of J_{nT} . This assumption is related to covariance stationarity but weaker. It is implied by covariance stationarity but it can hold even though covariance stationarity is violated. Kim and Sun (2011) give an example of a nonstationary spatial process that satisfies the above assumption. Stationarity seems to be a very strong assumption especially in the spatial dimension because a spatial process can be nonstationary simply because each unit has different numbers of neighbors. Assumption I6 is similar to the homogeneity assumption

in Bester, Conley and Hansen (2011). They assume that the covariance matrix in each cluster converges to the same limit.

The asymptotic bias of \tilde{J}_{nT} is determined by the smoothness of the kernels at zero and the decaying rates of the spatial and temporal dependence in terms of d_{ij} and d_{ts} . Define

$$K_q^{(a)} = \lim_{x \rightarrow 0} \frac{1 - K_a(x)}{|x|^q}, \quad \text{for } a = 1, 2 \text{ and } q \in [0, \infty).$$

and let $q_a = \max\{q : K_q^{(a)} < \infty\}$ be the *Parzen characteristic exponent* of $K_a(x)$ and $K_{q_a} = K_{q_a}^{(a)}$. The magnitude of q_a reflects the smoothness of $K_a(x)$ at $x = 0$.

Assumption I7 (i) The kernel functions $K_a(\cdot)$ with $a = 1, 2$ satisfy $K_a(0) = 1$, $|K_a(x)| \leq 1$, $K_a(x) = K_a(-x)$, $K_a(x) = 0$ for $|x| \geq 1$, and $q_1 \leq q_S$ and $q_2 \leq q_T$. (ii) For all $x_1, x_2 \in \mathbb{R}$ there is a constant, $c_L < 0$, such that

$$|K_a(x_1) - K_a(x_2)| \leq c_L |x_1 - x_2| \text{ for } a = 1, 2.$$

(iii) $\ell_{i,n}^{-1} \sum_{j=1}^n K_1^2\left(\frac{d_{ij}}{d_n}\right) \rightarrow \bar{K}_1$ for all $i \in E_n$.

Examples of kernels which satisfy Assumptions I7(i) and (ii) are the Bartlett, Tukey-Hanning and Parzen kernels. The quadratic spectral (QS) kernel does not satisfy Assumption I7(i) because it does not truncate. We may generalize our results to include the QS kernel but this requires much longer proofs. Assumption I7(iii) is more of an assumption on the distribution of the units. When the observations are located on a 2-dimensional integer lattice and d_{ij} is the Euclidian distance, we have

$$\bar{K}_1 = \frac{1}{\pi} \int_{-1}^1 \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} K_1^2\left(\sqrt{x^2+y^2}\right) dy dx = 2 \int_0^1 r K_1^2(r) dr.$$

In finite samples, we may use

$$\bar{K}_n = (n\ell_n)^{-1} \sum_{i,j=1}^n K_1^2\left(\frac{d_{ij}}{d_n}\right)$$

for \bar{K}_1 . Similarly we define, for $t \in E_T$,

$$\ell_{t,T}^{-1} \sum_{s=1}^T K_2^2\left(\frac{d_{ts}}{d_T}\right) \rightarrow \int_0^1 K_2^2(r) dr := \bar{K}_2.$$

Under Assumptions I7(i), we can define

$$b_1^{(q_1)} = \lim_{(n,T) \rightarrow \infty} b_n^{(q_1)}, \text{ where } b_n^{(q_1)} = \frac{1}{nT} \sum_{i,j=1}^n \sum_{t,s=1}^T \Gamma_{(it,js)} d_{ij}^{q_1},$$

$$b_2^{(q_2)} = \lim_{(n,T) \rightarrow \infty} b_T^{(q_2)}, \text{ where } b_T^{(q_2)} = \frac{1}{nT} \sum_{i,j=1}^n \sum_{t,s=1}^T \Gamma_{(it,js)} d_{ts}^{q_2}.$$

Let tr denote the trace function and \mathbb{K}_{pp} denote the $p^2 \times p^2$ commutation matrix. Under the assumptions above, we have the following theorem.

Theorem 1 Suppose that Assumptions 1 and I1-I7 hold, $d_n, d_T \rightarrow \infty$, $n_2 = o(n)$, $T_2 = o(T)$, $\ell_n = o(n)$ and $\ell_T = o(T)$.

- (a) $\lim_{(n,T) \rightarrow \infty} \frac{nT}{\ell_n \ell_T} \text{var}(\text{vec}(\tilde{J}_{nT})) = \bar{K}_1 \bar{K}_2 (I_{pp} + \mathbb{K}_{pp}) (J \otimes J)$.
- (b) If $d_n^{q_1}/d_T^{q_2} \rightarrow c_d$ as $(n, T) \rightarrow \infty$, then $\lim_{(n,T) \rightarrow \infty} d_n^{q_1} (E\tilde{J}_{nT} - J_{nT}) = -K_{q_1} b_1^{(q_1)} - c_d K_{q_2} b_2^{(q_2)}$.
- (c) If $d_n^{q_1}/d_T^{q_2} \rightarrow c_d$ and $d_n^{2q_1} \ell_n \ell_T / (nT) \rightarrow \tau \in (0, \infty)$, then

$$\begin{aligned} & \lim_{(n,T) \rightarrow \infty} AMSE \left(\frac{nT}{\ell_n \ell_T}, \hat{J}_{nT}, S \right) \\ &:= \lim_{(n,T) \rightarrow \infty} \frac{nT}{\ell_n \ell_T} E \left[\text{vec}(\tilde{J}_{nT} - J_{nT})' S \text{vec}(\tilde{J}_{nT} - J_{nT}) \right] \\ &= \frac{1}{\tau} \text{vec} \left(K_{q_1} b_1^{(q_1)} + c_d K_{q_2} b_2^{(q_2)} \right)' S \text{vec} \left(K_{q_1} b_1^{(q_1)} + c_d K_{q_2} b_2^{(q_2)} \right) \\ &+ \bar{K}_1 \bar{K}_2 \text{tr} (S(I_{pp} + \mathbb{K}_{pp})(J \otimes J)). \end{aligned}$$

Theorem 1(a) and (b) show that the asymptotic variance and bias of \tilde{J}_{nT} depend on the choice of d_n and d_T . When we increase d_n and/or d_T , the asymptotic bias decreases while the asymptotic variance increases. The convergence rate of \tilde{J}_{nT} is obtained by balancing the variance and the squared bias of \tilde{J}_{nT} . Accordingly, the rate of convergence of \tilde{J}_{nT} is $\sqrt{\ell_n \ell_T / (nT)}$. Under Assumption I5, the rate of convergence of \hat{J}_{nT} is also $\sqrt{\ell_n \ell_T / (nT)}$. If we set $\ell_n = O(d_n^{\eta_n})$ and $\ell_T = O(d_T^{\eta_T})$ for some $\eta_n > 0$ and $\eta_T = 1$, then the rate of convergence under the rate condition $d_n^{2q_1} \ell_n \ell_T / (nT) \rightarrow \tau \in (0, \infty)$ is $(nT)^{-q_1 q_2 / (q_1 \eta_T + 2q_1 q_2 + q_2 \eta_n)}$.

As \hat{J}_{nT} is consistent, the limiting distribution of the Wald statistic is the χ_g^2 distribution. This is a standard result. Under H_0 , $\mathcal{W}_{nT} \xrightarrow{d} \chi_g^2$ and $F_{nT} \xrightarrow{d} \chi_g^2 / g$.

4 Optimal bandwidth selection procedure

This section presents optimal bandwidth choice that minimizes an upper bound of AMSE of \hat{J}_{nT} and proposes a parametric plug-in procedure for practical implementation.

Let

$$B_{11} = \text{vec}(b_1^{(q_1)})' S \text{vec}(b_1^{(q_1)}), \quad B_{22} = \text{vec}(b_2^{(q_2)})' S \text{vec}(b_2^{(q_2)}), \quad B_{12} = \text{vec}(b_1^{(q_1)})' S \text{vec}(b_2^{(q_2)}).$$

Then up to smaller order terms the approximate MSE is

$$\begin{aligned} AMSE &:= AMSE \left(1, \hat{J}_{nT}, S \right) \\ &= \left(K_{q_1}^2 \frac{B_{11}}{d_n^{2q_1}} + 2K_{q_1} K_{q_2} \frac{B_{12}}{d_n^{q_1} d_T^{q_2}} + K_{q_2}^2 \frac{B_{22}}{d_T^{2q_2}} \right) + \frac{\ell_n \ell_T}{nT} \bar{K}_1 \bar{K}_2 \text{tr} [S (I_{pp} + \mathbb{K}_{pp}) (J \otimes J)] \\ &\leq 2 \left(K_{q_1}^2 \frac{B_{11}}{d_n^{2q_1}} + K_{q_2}^2 \frac{B_{22}}{d_T^{2q_2}} \right) + \frac{\ell_n \ell_T}{nT} \bar{K}_1 \bar{K}_2 \text{tr} [S (I_{pp} + \mathbb{K}_{pp}) (J \otimes J)] \\ &:= AMSE^*. \end{aligned}$$

AMSE* can be regarded as AMSE in the worst case:

$$AMSE^* = \max_{(b_1^{(q_1)}, b_2^{(q_2)}) \in \mathfrak{B}} AMSE,$$

where

$$\mathfrak{B} = \left\{ (b_1^{(q_1)}, b_2^{(q_2)}) : \text{vec}(b_1^{(q_1)})' S \text{vec}(b_1^{(q_1)}) = B_{11}, \text{vec}(b_2^{(q_2)})' S \text{vec}(b_2^{(q_2)}) = B_{22} \right\}.$$

Assuming $B_{11} > 0$ and $B_{22} > 0$, we select (d_n^*, d_T^*) to minimize the dominating terms in AMSE*:

$$(d_n^*, d_T^*) = \arg \min_{d_n, d_T} 2 \left(K_{q_1}^2 \frac{B_{11}}{d_n^{2q_1}} + K_{q_2}^2 \frac{B_{22}}{d_T^{2q_2}} \right) + \frac{\ell_n \ell_T}{nT} \bar{\kappa}_1 \bar{\kappa}_2 C_V, \quad (13)$$

where $C_V = \text{tr}[S(I_{pp} + \mathbb{K}_{pp})(J \otimes J)]$.

Here we use the AMSE* instead of the AMSE as the criterion, as the latter is intractable. The source of the problem is that B_{12} can be negative. In theory, we may choose d_n and d_T to zero out the bias terms under some conditions. For example, consider the case $B_{12} = -\sqrt{B_{11}B_{22}}$. This may occur when we are interested in a single component of β . In this case, bandwidth parameters satisfying $d_n^{q_1}/d_T^{q_2} = (K_{q_1}/K_{q_2}) \sqrt{B_{11}/B_{22}}$ make the first order bias terms cancel out with each other. Therefore, in theory, we need to select d_n or d_T to tradeoff the second-order bias with the variance. However, this choice is infeasible in practice. As B_{11}/B_{22} is unknown, we have to estimate this ratio and the estimation error is of the same order as the first order bias. So the first order bias cannot be reduced by an order of magnitude in practice. Our minimax criterion avoids this problem. It is also simple to implement, as d_n^* and d_T^* depend only on two bias terms but not on their interaction B_{12} . It also effectively controls for the AMSE in terms of an upper bound, which is achievable under some data generating processes.

Under the boundary condition in the time dimension, we have $\ell_T/T \rightarrow 0$, $\ell_T = 2d_T + o(d_T)$. In some cases, it is also possible to approximate ℓ_n as a function of d_n . For example, if individuals are located on a 2-dimensional lattice and the Euclidean distance is used, $\ell_n = \pi d_n^2$ would be a reasonable approximation. With the specification of $\ell_n = \alpha_n d_n^{\eta_n}$ and $\ell_T = \alpha_T d_T^{\eta_T}$, we obtain explicit formulae of d_n^* and d_T^* as follows:

$$d_n^* = \left(\frac{4q_1 K_{q_1}^2 B_{11}}{\eta_n \alpha_n \alpha_T \bar{\kappa}_1 \bar{\kappa}_2 C_V} nT \right)^{\frac{q_2}{q_1 \eta_T + 2q_1 q_2 + q_2 \eta_n}} \left(\frac{q_1 K_{q_1}^2 \eta_T B_{11}}{q_2 K_{q_2}^2 \eta_n B_{22}} \right)^{\frac{\eta_T}{2(q_1 \eta_T + 2q_1 q_2 + q_2 \eta_n)}}, \quad (14)$$

$$d_T^* = \left(\frac{4q_2 K_{q_2}^2 B_{22}}{\eta_T \alpha_n \alpha_T \bar{\kappa}_1 \bar{\kappa}_2 C_V} nT \right)^{\frac{q_1}{q_1 \eta_T + 2q_1 q_2 + q_2 \eta_n}} \left(\frac{q_2 K_{q_2}^2 \eta_n B_{22}}{q_1 K_{q_1}^2 \eta_T B_{11}} \right)^{\frac{\eta_n}{2(q_1 \eta_T + 2q_1 q_2 + q_2 \eta_n)}}. \quad (15)$$

The optimal bandwidth formulae in (14) and (15) show that the degree of persistence in one dimension affects both d_n^* and d_T^* but in opposite directions. For example, if a process becomes spatially persistent, d_n^* is increased to address the increasing bias, which comes from the usage of kernel truncation in the spatial domain. But, the increase of d_n^* , at the same time, magnifies the variance term. Therefore, in order to minimize the AMSE*, d_T^* is decreased to moderate the inflation of the asymptotic variance. Figure 1 illustrates this relation of d_n^* and d_T^* with different dependence structure. The two graphs are the level curves of d_n^* and d_T^* as functions of λ and ρ , which determine the temporal and spatial persistence respectively in the following DGP:

$$V_t = \lambda V_{t-1} + u_t, \quad u_t = \rho W_n u_t + \varepsilon_t \text{ and } \varepsilon_t \sim (0, I_n),$$

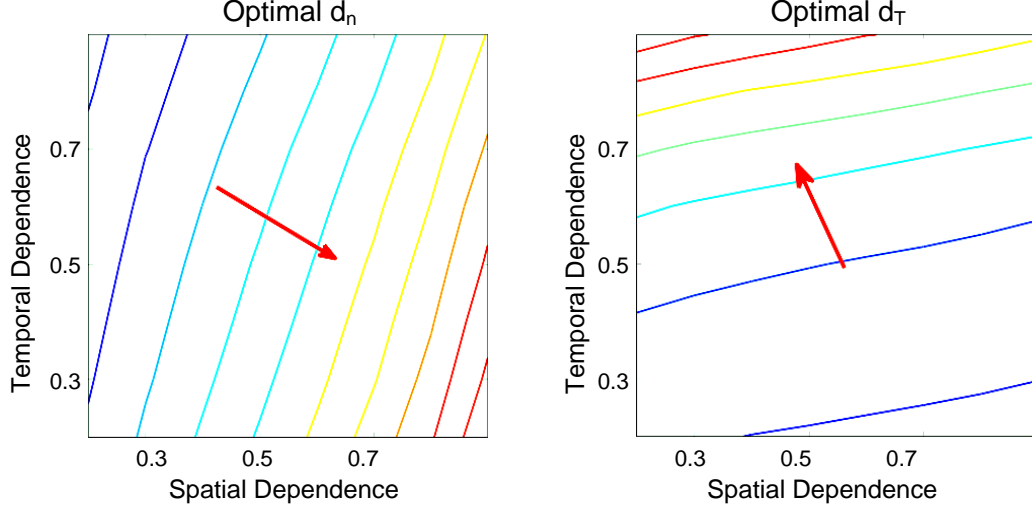


Figure 1: Level curves of d_n^* and d_T^* as functions of spatial and temporal dependence

where V_t , u_t and ε_t are n -vectors such as $V_t = (V_{(1,t)}, V_{(2,t)}, \dots, V_{(n,t)})'$ and W_n is a spatial weight matrix. These two graphs indicate that d_n^* increases as spatial dependence increases or temporal dependence decreases and that d_T^* increases as temporal dependence grows or spatial dependence is reduced.

The corollary below gives a precise sense that (d_n^*, d_T^*) is optimal.

Corollary 1 *Suppose Assumptions 1 and I1-I7 hold. Assume that $\ell_n = \alpha_n d_n^{\eta_n}$ and $\ell_T = \alpha_T d_T^{\eta_T}$ for some $\eta_n, \eta_T > 0$, $\alpha_n = \alpha_1 + o(1)$, $\alpha_T = \alpha_2 + o(1)$, $B_{11} > 0$ and $B_{22} > 0$. Then, for any sequence of bandwidth parameters $\{d_n, d_T\}$ such that $d_n^{q_1}/d_T^{q_2} \rightarrow c_d \in (0, \infty)$ and $d_n^{2q_1} \ell_n \ell_T / (nT) \rightarrow \tau \in (0, \infty)$, $\{d_n^*, d_T^*\}$ is preferred in the sense that*

$$\lim_{(n,T) \rightarrow \infty} \left[\max_{(b_1^{(q_1)}, b_2^{(q_2)}) \in \mathfrak{B}} \text{AMSE} \left((nT)^{2q_1 q_2 / (q_1 \eta_T + 2q_1 q_2 + q_2 \eta_n)}, \hat{J}_{nT}(d_n, d_T), S \right) - \max_{(b_1^{(q_1)}, b_2^{(q_2)}) \in \mathfrak{B}} \text{AMSE} \left((nT)^{2q_1 q_2 / (q_1 \eta_T + 2q_1 q_2 + q_2 \eta_n)}, \hat{J}_{nT}(d_n^*, d_T^*), S \right) \right] \geq 0.$$

The inequality is strict unless $d_n = d_n^* (1 + o(1))$ and $d_T = d_T^* (1 + o(1))$.

Theorem 1 and Corollary 1 are applicable only to finite order kernels. This rules out the flat-top kernels which are infinite order kernels from a frequency domain perspective. In their general form, the class of flat top kernels is given by

$$\mathfrak{K}_F = \left(\mathfrak{K}(\cdot) : \mathfrak{K}(x) = \begin{cases} 1 & \text{if } |x| \leq c_F \\ \mathcal{G}(x) & \text{otherwise} \end{cases} \right) \quad (16)$$

where $c_F \leq 1$ and $\mathcal{G} : |x| \in (c_F, 1] \rightarrow [0, 1]$. A typical flat-top kernel in \mathfrak{K}_F is the trapezoidal kernel in which $\mathcal{G}(x) = \max\{(|x| - 1)/(c_F - 1), 0\}$. The rectangular kernel is an extreme case with $c_F = 1$. For a flat-top kernel covariance estimator, the asymptotic bias is of smaller order than

that in Theorem 1(b). As a result, our bandwidth selection procedure does not apply directly to flat-top kernel estimators. However, it is interesting to consider flat-top kernel estimators because they are higher-order accurate. This is particularly important in our setting because flat-top kernels are completely compatible with the adaptiveness of our estimator as explained below while finite-order kernels yield some discrepancy. In time series HAC estimation, Andrews (1991, footnote on p.834) and Lin and Sakata (2012) suggest a practical bandwidth rule for the rectangular kernel estimator, a special flat-top kernel estimator, based on the MSE criterion. Sun and Kaplan (2011) explore this problem rigorously and provide a bandwidth selection procedure that is testing optimal. We extend these methods to the present setting. For any finite-order kernel estimator set as the target, we can select the bandwidth parameters for the flat-top kernel $(d_{F,n}^*, d_{F,T}^*)$ such that the flat-top-kernel-based covariance estimator has a smaller AMSE*.

Let $K_{tar,1}(\cdot)$ and $K_{tar,2}(\cdot)$ be the target kernels in the spatial and time domains and $(d_{tar,n}^*, d_{tar,T}^*)$ be the respective optimal bandwidth parameters. Let $K_{F,1}(\cdot)$ and $K_{F,2}(\cdot)$ be the flat-top kernels used in the two domains. Given $\ell_n = \alpha_n d_n^{\eta_n}$ and $\ell_T = \alpha_T d_T^{\eta_T}$, if we set

$$d_{F,n}^* = d_{tar,n}^* \left(\frac{\bar{K}_{tar,1}}{\bar{K}_{F,1}} \right)^{1/\eta_n} \quad \text{and} \quad d_{F,T}^* = d_{tar,T}^* \left(\frac{\bar{K}_{tar,2}}{\bar{K}_{F,2}} \right)^{1/\eta_T}, \quad (17)$$

then the asymptotic variance of the flat-top-kernel estimator is the same as that of the estimator based on the target kernel. However, under some smoothness conditions, the asymptotic bias of the flat-top kernel estimator is of smaller order. As a result, the flat-top kernel estimator has smaller AMSE* than that based on the target kernel.

The unknown values such as B_{11} , B_{22} and C_V in the optimal bandwidth formula (13) can be estimated using a parametric plug-in method (e.g. Andrews, 1991; and Kim and Sun, 2011). We consider the following four different spatiotemporal parametric models, which are introduced in Anselin (2001):

$$V_{(i,t)}^{(c)} = \rho_c \left[W_n^{(c)} V_{t-1}^{(c)} \right]_i + \tilde{\varepsilon}_{(i,t)}^{(c)}, \quad (18)$$

$$V_{(i,t)}^{(c)} = \lambda_c V_{(i,t-1)}^{(c)} + \rho_c \left[W_n^{(c)} V_{t-1}^{(c)} \right]_i + \tilde{\varepsilon}_{(i,t)}^{(c)} \quad (19)$$

$$V_{(i,t)}^{(c)} = \lambda_c V_{(i,t-1)}^{(c)} + \phi_c \left[W_n^{(c)} V_t^{(c)} \right]_i + \tilde{\varepsilon}_{(i,t)}^{(c)} \quad (20)$$

$$V_{(i,t)}^{(c)} = \lambda_c V_{(i,t-1)}^{(c)} + \phi_c \left[W_n^{(c)} V_t^{(c)} \right]_i + \rho_c \left[W_n^{(c)} V_{t-1}^{(c)} \right]_i + \tilde{\varepsilon}_{(i,t)}^{(c)} \quad (21)$$

where $\tilde{\varepsilon}_{(i,t)}^{(c)} \stackrel{i.i.d}{\sim} (0, \sigma_{cc})$ and $[W_n^{(c)} V_t^{(c)}]_i$ is the i^{th} element of vector $W_n^{(c)} V_t^{(c)}$. The spatial weight matrix $W_n^{(c)}$ is determined *a priori* and by convention it is row-standardized and its diagonal elements are zeros.⁷

For an illustrative purpose, consider the model in (18). It can be rewritten recursively as

⁷The way to construct a spatial weight matrix is well explained in the spatial econometrics literature (e.g. LeSage and Pace, 2009).

follows:

$$\begin{aligned}
V_1^{(c)} &= \rho_c W_n^{(c)} V_0^{(c)} + I_n \tilde{\varepsilon}_1^{(c)} \\
V_2^{(c)} &= \rho_c^2 \left(W_n^{(c)}\right)^2 V_0^{(c)} + \rho_c W_n^{(c)} \tilde{\varepsilon}_1^{(c)} + I_n \tilde{\varepsilon}_2^{(c)} \\
&\vdots \\
V_T^{(c)} &= \rho_c^T \left(W_n^{(c)}\right)^T V_0^{(c)} + \rho_c^{T-1} \left(W_n^{(c)}\right)^{T-1} \tilde{\varepsilon}_1^{(c)} + \rho_c^{T-2} \left(W_n^{(c)}\right)^{T-2} \tilde{\varepsilon}_2^{(c)} + \dots + I_n \tilde{\varepsilon}_T^{(c)}
\end{aligned}$$

Imposing the initial condition of $V_0 = 0$, we can estimator ρ_c by OLS with $\hat{V}_t^{(c)} = (\hat{V}_{(1,t)}^{(c)}, \dots, \hat{V}_{(n,t)}^{(c)})'$. We define

$$\hat{R}_{ts}^{(c)} = \begin{cases} I_n, & \text{if } t - s = 0 \\ \left(\hat{\rho}_c W_n^{(c)}\right)^{t-s}, & \text{if } t - s > 0 \\ 0, & \text{otherwise,} \end{cases}$$

and

$$\hat{R}_{(i,t)}^{(c)} = \left[\hat{R}_{t1,i}^{(c)}, \hat{R}_{t2,i}^{(c)}, \dots, \hat{R}_{tT,i}^{(c)} \right],$$

where $\hat{R}_{ts,i}^{(c)}$ denotes the i -th row of $\hat{R}_{ts}^{(c)}$. Consequently, we approximate J , $b_1^{(q_1)}$ and $b_2^{(q_2)}$ by

$$\hat{J}(c, d) = \frac{\hat{\sigma}_{cd}}{nT} \sum_{i,j=1}^n \sum_{t,s=1}^T \left(\hat{R}_{(i,t)}^{(c)} \right) \left(\hat{R}_{(j,s)}^{(d)} \right)', \quad (22)$$

$$\hat{b}_1^{(q_1)}(c, d) = \frac{\hat{\sigma}_{cd}}{nT} \sum_{i,j=1}^n \sum_{t,s=1}^T \left(\hat{R}_{(i,t)}^{(c)} \right) \left(\hat{R}_{(j,s)}^{(d)} \right)' d_{ij}^{q_1}, \quad (23)$$

$$\hat{b}_2^{(q_2)}(c, d) = \frac{\hat{\sigma}_{cd}}{nT} \sum_{i,j=1}^n \sum_{t,s=1}^T \left(\hat{R}_{(i,t)}^{(c)} \right) \left(\hat{R}_{(j,s)}^{(d)} \right)' d_{ts}^{q_2}, \quad (24)$$

where

$$\hat{\sigma}_{cd} = \frac{1}{n(T-1)-1} \left(\hat{\varepsilon}^{(c)} \right)' \left(\hat{\varepsilon}^{(d)} \right),$$

$\hat{\varepsilon}^{(c)} = ((\hat{\varepsilon}_1^{(c)})', \dots, (\hat{\varepsilon}_T^{(c)})')'$, $\hat{\varepsilon}_1^{(c)} = \hat{V}_1^{(c)}$ and $\hat{\varepsilon}_t^{(c)} = \hat{V}_t^{(c)} - \hat{\rho}_c W_n^{(c)} \hat{V}_{t-1}^{(c)}$ for $t \geq 2$. Substituting these estimates into the optimal bandwidth formulae, we obtain the data-driven bandwidth parameters (\hat{d}_n, \hat{d}_T) as follows:

$$\hat{d}_n = \left(\frac{4q_1 K_{q_1}^2 \hat{B}_{11}}{\eta_n \alpha_n \alpha_T \bar{\mathcal{K}}_1 \bar{\mathcal{K}}_2 \hat{C}_V} nT \right)^{\frac{q_2}{q_1 \eta_T + 2q_1 q_2 + q_2 \eta_n}} \left(\frac{q_1 K_{q_1}^2 \eta_T \hat{B}_{11}}{q_2 K_{q_2}^2 \eta_n \hat{B}_{22}} \right)^{\frac{\eta_T}{2(q_1 \eta_T + 2q_1 q_2 + q_2 \eta_n)}}, \quad (25)$$

$$\hat{d}_T = \left(\frac{4q_2 K_{q_2}^2 \hat{B}_{22}}{\eta_T \alpha_n \alpha_T \bar{\mathcal{K}}_1 \bar{\mathcal{K}}_2 \hat{C}_V} nT \right)^{\frac{q_1}{q_1 \eta_T + 2q_1 q_2 + q_2 \eta_n}} \left(\frac{q_2 K_{q_2}^2 \eta_n \hat{B}_{22}}{q_1 K_{q_1}^2 \eta_T \hat{B}_{11}} \right)^{\frac{\eta_n}{2(q_1 \eta_T + 2q_1 q_2 + q_2 \eta_n)}}. \quad (26)$$

where

$$\hat{B}_{11} = \text{vec} \left(\hat{b}_1^{(q_1)} \right)' \text{Svec} \left(\hat{b}_1^{(q_1)} \right), \quad \hat{B}_{22} = \text{vec} \left(\hat{b}_2^{(q_2)} \right)' \text{Svec} \left(\hat{b}_2^{(q_2)} \right), \quad \hat{C}_V = \text{tr} \left[S(I_{pp} + \mathbb{K}_{pp})(\hat{J} \otimes \hat{J}) \right].$$

Since the models in (19), (20) and (21) can be rewritten as

$$\begin{aligned} V_{(i,t)}^{(c)} &= \left[\left(\lambda_c I_n + \rho_c W_n^{(c)} \right) V_{t-1}^{(c)} \right]_i + \tilde{\varepsilon}_{it}^{(c)}, \\ V_{(i,t)}^{(c)} &= \left[\lambda_c \left(I_n - \phi_c W_n^{(c)} \right)^{-1} V_{t-1}^{(c)} \right]_i + \left[\left(I_n - \phi_c W_n^{(c)} \right)^{-1} \tilde{\varepsilon}_t^{(c)} \right]_i, \\ V_{(i,t)}^{(c)} &= \left[\left(I_n - \phi_c W_n^{(c)} \right)^{-1} \left(\lambda_c I_n + \rho_c W_n^{(c)} \right) V_{t-1}^{(c)} \right]_i + \left[\left(I_n - \phi_c W_n^{(c)} \right)^{-1} \tilde{\varepsilon}_t^{(c)} \right]_i, \end{aligned}$$

we can derive the data-driven bandwidth parameters with these models using the same procedures as (18). While the OLS estimator is consistent for (19), it is not for (20) and (21) due to the endogeneity of $[W_n^{(c)} V_t^{(c)}]_i$. For these models, we can obtain consistent estimators using QMLE as follows:

$$\left(\hat{\lambda}_c, \hat{\phi}_c, \hat{\rho}_c, \hat{\sigma}_{cc} \right) = \arg \min_{\lambda_c, \phi_c, \rho_c, \sigma_{cc}} \frac{1}{2} \ln \sigma_{cc} - \frac{1}{n} \ln \left| I_n - \phi_c W_n^{(c)} \right| + \frac{1}{2\sigma_{cc}} \frac{1}{nT} \sum_{t=1}^T \left(\hat{\varepsilon}_t^{(c)} \right)' \left(\hat{\varepsilon}_t^{(c)} \right).$$

See Yu, de Jong and Lee (2008) for details. However, we argue that the simple OLS can still be used for (20) and (21). Since the parametric models are likely to be mis-specified, the QML estimator is not necessarily preferred. In addition, as argued by Andrews (1991), good performance of the estimator only requires (\hat{d}_n, \hat{d}_T) to be near the optimal bandwidth values and not to be precisely equal to them. Furthermore, OLS estimation is computationally much less demanding.

5 Comparison with CCE, DK and DK* estimators

For comparison, we examine the asymptotic properties of the CCE, DK and DK* estimators based on our data representation in (8) and (9) under the increasing-smoothing asymptotics. We also derive the optimal bandwidth parameters for the DK and DK* estimators using the AMSE criterion.

5.1 CCE

The CCE is defined as

$$\hat{J}_{nT}^A = \frac{1}{nT} \sum_{i=1}^n \sum_{t,s=1}^T \hat{V}_{(i,t)} \hat{V}_{(i,s)}'.$$

Define \tilde{J}_{nT}^A in the same way but with $\hat{V}_{(i,t)}$ replaced by $V_{(i,t)}$. The crucial condition for \hat{J}_{nT}^A to be consistent is that variables for two different individuals (or clusters) are uncorrelated, i.e. $EV_{(i,t)} V_{(j,s)}' = 0$ if $i \neq j$. Under this condition, \hat{J}_{nT}^A is robust to heteroskedasticity and arbitrary forms of time series correlation. Our spatiotemporal representation accommodates spatial independence by imposing the following restriction.

Assumption I8 $\tilde{r}_{(it,j,s)} = 0$ if $i \neq j$.

Under Assumption I8, we have

$$J_{nT} = \frac{1}{nT} \sum_{i=1}^n \sum_{t,s=1}^T E \left[V_{(i,t)} V_{(i,s)}' \right].$$

Assumption I9 For all $i \in E_n$,

$$\lim_{T \rightarrow \infty} \text{var} \left(\frac{1}{\sqrt{T}} \sum_{s=1}^T V_{(i,s)} \right) = J.$$

Assumption I9 implies the homogeneity of $\text{var}(T^{-1/2} \sum_{s=1}^T V_{(i,s)})$ for $i \in E_n$, under which we can derive the asymptotic variance of $\tilde{J}_{nT}^A = (nT)^{-1} \sum_{i=1}^n \sum_{t,s=1}^T V_{(i,t)} V'_{(i,s)}$ in Theorem 2 below.

Theorem 2 Suppose that Assumptions 1, I1, I2, I8 and I9 hold. Then $\lim_{(n,T) \rightarrow \infty} n\text{var}(\text{vec}(\tilde{J}_{nT}^A)) = (I_{pp} + \mathbb{K}_{pp})(J \otimes J)$.

The proof is analogous to the proof of Theorem 1(a) and is omitted here for brevity.⁸ Theorem 2 and the fact that $E\tilde{J}_{nT}^A = J_{nT}$ imply the \sqrt{n} -convergence of \tilde{J}_{nT}^A . Under the sufficient conditions for Assumption I5 given in the supplementary appendix, we have $\hat{J}_{nT}^A - \tilde{J}_{nT}^A = o_p(1/\sqrt{n})$. Hence \hat{J}_{nT}^A also converges to J at the rate of $1/\sqrt{n}$, which is consistent with Hansen (2007).

5.2 DK estimator

The DK estimator is based on the time series HAC estimation method with cross sectional averages. The estimator is defined as

$$\hat{J}_{nT}^{DK} = \frac{1}{nT} \sum_{i,j=1}^n \sum_{t,s=1}^T K_2 \left(\frac{d_{ts}}{d_T} \right) \hat{V}_{(i,t)} \hat{V}'_{(j,s)}.$$

Similarly, we define \tilde{J}_{nT}^{DK} as above but with $\hat{V}_{(i,t)}$ replaced by $V_{(i,t)}$.

For the asymptotic properties, we introduce the following assumptions in place of Assumption I6.

Assumption I10 As $d_T \rightarrow \infty$ with $(n, T) \rightarrow \infty$, given $\ell_T = o(T)$,

$$\lim_{(n,T) \rightarrow \infty} \text{var} \left(\frac{1}{\sqrt{n\ell_{t,T}}} \sum_{j=1}^n \sum_{s: d_{ts} \leq d_T} V_{(j,s)} \right) = J,$$

for all $t \in E_T$.

Theorem 3 below gives the asymptotic properties of \tilde{J}_{nT}^{DK} and \hat{J}_T^{DK} . Its proof is omitted here as it is similar to the proof of Theorem 1.

Theorem 3 Suppose that Assumptions 1, I1, I2, I3(ii), I5, I7(i)(ii), and I10 hold, and $d_T \rightarrow \infty$, $\ell_T = o(T)$.

$$(a) \lim_{(n,T) \rightarrow \infty} \frac{T}{\ell_T} \text{var} \left(\text{vec}(\tilde{J}_{nT}^{DK}) \right) = \bar{K}_2(I_{pp} + \mathbb{K}_{pp})(J \otimes J).$$

$$(b) \lim_{(n,T) \rightarrow \infty} d_T^{q_2} (E\tilde{J}_{nT}^{DK} - J_{nT}) = -K_{q_2} b_2^{(q_2)}.$$

⁸Detailed proofs for Theorems 2–4 are available from the authors upon request.

(c) If $d_T^{2q_2} \ell_T / T \rightarrow \tau \in (0, \infty)$, then

$$\begin{aligned} & \lim_{(n,T) \rightarrow \infty} AMSE \left(\frac{T}{\ell_T}, \hat{J}_{nT}^{DK}, S \right) \\ &= \frac{1}{\tau} K_{q_2}^2 \left(\text{vec } b_2^{(q_2)} \right)' S \left(\text{vec } b_2^{(q_2)} \right) + \bar{K}_2 \text{tr} [S(I_{pp} + \mathbb{K}_{pp})(J \otimes J)]. \end{aligned}$$

Theorem 3(a) and (b) imply that \tilde{J}_{nT}^{DK} is consistent if $d_T \rightarrow \infty$ and $\ell_T = o(T)$. The rate of convergence obtained by balancing the variance and the squared bias is $T^{-q_2/(2q_2+\eta_T)}$. This is also the rate of convergence for \hat{J}_{nT}^{DK} under the sufficient conditions for Assumption I5 given in the supplementary appendix. Therefore, the rate of convergence of \hat{J}_{nT} is faster than that of \hat{J}_{nT}^{DK} if $T = o(n^{q_1(2q_2+\eta_T)/(q_2\eta_n)})$.

The optimal bandwidth parameter of \hat{J}_{nT}^{DK} based on the AMSE criterion is

$$d_T^{DK} = \left(\frac{2q_2 K_{q_2}^2 B_{22}}{\eta_T \alpha_T \bar{K}_2 C_V} T \right)^{1/(2q_2+\eta_T)}. \quad (27)$$

We can obtain the data-driven bandwidth parameter following Andrews (1991).

5.3 DK* estimator

Analogous to the DK estimator, we can consider the usage of spatial HAC estimation applied to time series averages, especially when n is relatively large and T is relatively small. The DK* estimator based on the time series averages is

$$\hat{J}_{nT}^{DK*} = \frac{1}{nT} \sum_{i,j=1}^n \sum_{t,s=1}^T K_1 \left(\frac{d_{ij}}{d_n} \right) \hat{V}_{(i,t)} \hat{V}_{(j,s)}'.$$

Let \tilde{J}_{nT}^{DK*} denote the infeasible version of \hat{J}_{nT}^{DK*} with $\hat{V}_{(i,t)}$ replaced by $V_{(i,t)}$.

Assumption I11 As $d_n \rightarrow \infty$ with $(n, T) \rightarrow \infty$, given $\ell_n = o(n)$,

$$\lim_{(n,T) \rightarrow \infty} \text{var} \left(\frac{1}{\sqrt{\ell_{i,nT}}} \sum_{j:d_{ij} \leq d_n} \sum_{s=1}^T V_{(j,s)} \right) = J,$$

for all $i \in E_n$.

Theorem 4 below gives the asymptotic properties of \tilde{J}_{nT}^{DK*} and \hat{J}_{nT}^{DK*} . The proof is similar to the proof of Theorem 1 and is omitted to save space.

Theorem 4 Suppose that Assumptions 1, I1, I2, I3(i), I4, I5, I7, and I11 hold, $n_2 = o(n)$, $\ell_n, d_n \rightarrow \infty$ and $\ell_n = o(n)$.

(a) $\lim_{(n,T) \rightarrow \infty} \frac{n}{\ell_n} \text{var} \left(\text{vec}(\tilde{J}_{nT}^{DK*}) \right) = \bar{K}_1(I_{pp} + \mathbb{K}_{pp})(J \otimes J).$

(b) $\lim_{(n,T) \rightarrow \infty} d_n^{q_1} (E \tilde{J}_{nT}^{DK*} - J_{nT}) = -K_{q_1} b_1^{(q_1)}.$

(c) If $d_n^{2q_1} \ell_n/n \rightarrow \tau \in (0, \infty)$, then

$$\begin{aligned} & \lim_{(n,T) \rightarrow \infty} AMSE \left(\frac{n}{\ell_n}, \hat{J}_{nT}^{DK*}, S \right) \\ &= \frac{1}{\tau} K_{q_1}^2 \text{vec} \left(b_1^{(q_1)} \right)' \text{Svec} \left(b_1^{(q_1)} \right) + \bar{\mathcal{K}}_1 \text{tr} [S(I_{pp} + \mathbb{K}_{pp})(J \otimes J)]. \end{aligned}$$

If we can characterize $\ell_n = \alpha_n d_n^{\eta_n}$, \hat{J}_{nT} achieves the faster convergence rate than \tilde{J}_{nT}^{DK*} and \hat{J}_{nT}^{DK*} if $n = o(T^{q_2(2q_1+\eta_n)/(q_1\eta_T)})$. The optimal bandwidth based on the AMSE criterion is

$$d_n^{DK*} = \left(\frac{2q_1 K_{q_1}^2 B_{11}}{\eta_n \alpha_n \bar{\mathcal{K}}_1 C_V} n \right)^{1/(2q_1+\eta_n)}. \quad (28)$$

We can obtain the data-driven bandwidth parameter following Kim and Sun (2011).

6 Adaptiveness of \hat{J}_{nT}

6.1 Flexibility

\hat{J}_{nT} is flexible in the sense that it includes the estimators in the previous section as special cases, reducing to each of them in large samples with certain choice of the bandwidths and kernel function.

In order to illustrate the flexibility, we introduce the generalized CCE estimator:

$$\hat{J}_{nT}^{GA} = \frac{1}{nT} \sum_{i=1}^n \sum_{t,s=1}^T K_2 \left(\frac{d_{ts}}{d_T} \right) \hat{V}_{(i,t)} \hat{V}_{(i,s)}'$$

with \tilde{J}_{nT}^{GA} as its infeasible version. \hat{J}_{nT}^{GA} includes \hat{J}_{nT}^A as a special case with $K_F(\cdot) \in \mathfrak{K}_F$ with $c_F = 1$ and $d_T = T$.

The following proposition shows the asymptotic equivalence of \hat{J}_{nT} to the existing estimators with certain sequences of d_n and d_T .

Proposition 1 *Let Assumptions I1 and I2 hold. Assume that $\hat{J}_{nT} = \tilde{J}_{nT} + o_p(1)$, $\hat{J}_{nT}^{GA} = \tilde{J}_{nT}^{GA} + o_p(1)$, $\hat{J}_{nT}^{DK} = \tilde{J}_{nT}^{DK} + o_p(1)$ and $\hat{J}_{nT}^{DK*} = \tilde{J}_{nT}^{DK*} + o_p(1)$.*

- (a) *If $\min_{i,j} (d_{ij}) > \varepsilon$ for all $i \neq j$ and some $\varepsilon > 0$ and $d_n \rightarrow 0$ as $n \rightarrow \infty$, then $\hat{J}_{nT} - \hat{J}_{nT}^{GA} = o_p(1)$.*
- (b) *If $K_1(\cdot)$ is the rectangular kernel, Assumption I3(i) holds, and $\ell_n/n \rightarrow 1$ as $n \rightarrow \infty$, then $\hat{J}_{nT} - \hat{J}_{nT}^{DK} = o_p(1)$.*
- (c) *If $K_2(\cdot)$ is the rectangular kernel, Assumption I3(ii) holds, and $\ell_T/T \rightarrow 1$ as $T \rightarrow \infty$, then $\hat{J}_{nT} - \hat{J}_{nT}^{DK*} = o_p(1)$.*

The flexibility of our estimator relies on the property that the rectangular kernel does not downweigh the covariances between spatially or temporally remote units. In contrast, \hat{J}_{nT} with finite-order kernels does not completely reduce to \hat{J}_{nT}^{DK} and \hat{J}_{nT}^{DK*} with large d_n and d_T , getting close to them though.

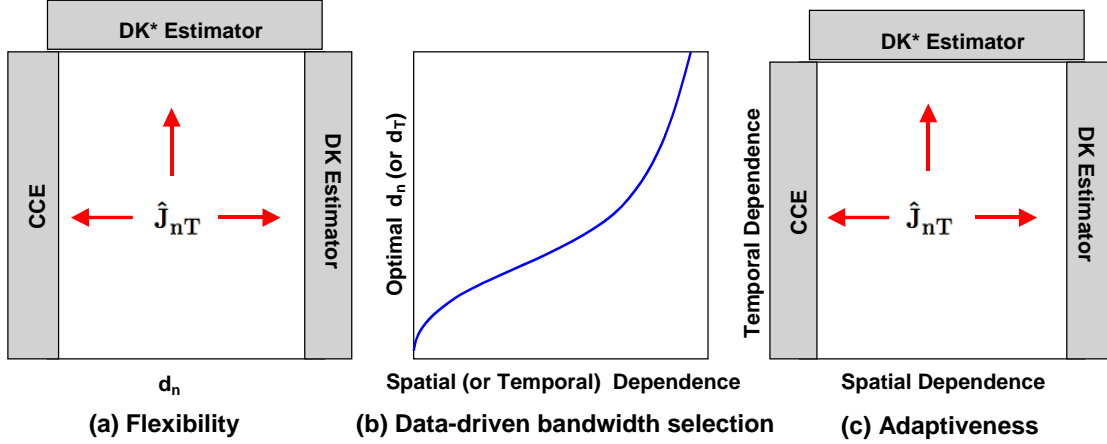


Figure 2: Adaptiveness of \hat{J}_{nT}

6.2 Adaptiveness

While \hat{J}_{nT} has advantages in terms of robustness over \hat{J}_{nT}^A and in terms of efficiency over \hat{J}_{nT}^{DK*} and \hat{J}_{nT}^{DK} , for certain dependence structure, one of the existing estimators is expected to outperform the other estimators. If a process is spatially highly persistent, \hat{J}_{nT}^{DK} is expected to outperform the other estimators in that it is robust to arbitrary forms of spatial correlation. For the same reason, \hat{J}_{nT}^{DK*} tends to perform better than the others, if a process is temporally highly persistent. \hat{J}_{nT}^A is more efficient than the other estimators in the absence of spatial correlation.

The attractiveness of our estimator \hat{J}_{nT} is that, with the data-driven bandwidth choice, it becomes close to the estimator that is expected to perform the best. This adaptiveness is the novel feature of our estimation method. It practically automates the selection of covariance estimators. As illustrated in Figure 2, adaptiveness arises from the flexibility and data-driven bandwidth selection procedure. In case that a process is spatially highly persistent, the data-driven bandwidth selection procedure yields large \hat{d}_n so that \hat{J}_{nT} gets close to \hat{J}_{nT}^{DK} . Analogously, \hat{J}_{nT} becomes close to \hat{J}_{nT}^{DK*} if a process is very persistent in the time dimension. In the absence of spatial dependence, \hat{J}_{nT} becomes close to \hat{J}_{nT}^{GA} with small \hat{d}_n .

It should be pointed out that finite-order kernels do not achieve complete adaptiveness because downweighing restricts its flexibility in bridging the existing estimators. We can fix this by employing a rectangular kernel. In this case, with appropriate bandwidth choices, \hat{J}_{nT} is asymptotically equivalent to the best estimator. The bandwidth selection rule in (17) meets the requirement, as the selected bandwidths from (17) are proportional to those from (13).⁹

⁹Another issue with flat-top kernel estimators is that they are not positive semi-definite. Politis (2011) and Lin and Sakata (2009) propose simple modifications to the estimator to enforce the positive (semi) definiteness without sacrificing asymptotic efficiency. In our simulation, we use the method suggested by Politis (2011).

7 Fixed-smoothing asymptotics

7.1 Limiting theory for \hat{J}_{nT} under fixed-smoothing asymptotics

Following Conley (1999) and Sun and Kim (2013), we assume that, given a distance measure, it is possible to map the individuals onto a 2-dimensional integer lattice so that d_{ij} can be expressed in terms of the lattice indices. Let the locations be indexed by $i = (i_1, i_2) = [1, 2, \dots, L_n] \otimes [1, 2, \dots, M_n]$. We can then rewrite the sample moment conditions that define $\hat{\beta}$ as

$$\frac{1}{nT} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^T 1_{i_1, i_2} \hat{V}_{(i_1, i_2, t)} = 0,$$

where $\hat{V}_{(i_1, i_2, t)}$ is associated with an observation located at (i_1, i_2) and time t , and the indicator function $1_{i_1, i_2}$ indicates whether an observation is present at the lattice point (i_1, i_2) .

We introduce the following assumption on the distance measure in the spatial dimension.

Assumption F1 Let $d_{(i_1, i_2), (j_1, j_2)}$ denote the distance between the two units located at (i_1, i_2) and (j_1, j_2) . Then,

$$\frac{d_{(i_1, i_2), (j_1, j_2)}}{d_n} = d\left(\frac{|i_1 - j_1|}{d_n}, \frac{|i_2 - j_2|}{d_n}\right)$$

and $d(\cdot, \cdot)$ is continuously differentiable.

Assumption F1 implies that $d_{(i_1, i_2), (j_1, j_2)}$ is a function of $|i_1 - j_1|$ and $|i_2 - j_2|$ and is homogeneous. This is not overly restrictive. p -norm distances that are usually employed in practice satisfy this assumption.

Let $b_1 = d_n/L_n$, $b_2 = d_n/M_n$ and $b_3 = d_T/T$. Suppose that the level of smoothing is held fixed such that $b = (b_1, b_2, b_3)$ are fixed constants. Then

$$\hat{J}_{nT} := \frac{1}{nT} \sum_{i_1, j_1=1}^{L_n} \sum_{i_2, j_2=1}^{M_n} \sum_{t, s=1}^T \mathbb{K}_b\left(\left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T}\right), \left(\frac{j_1}{L_n}, \frac{j_2}{M_n}, \frac{s}{T}\right)\right) \hat{V}_{(i_1, i_2, t)}^* \hat{V}_{(j_1, j_2, s)}^{*'}.$$

where $\hat{V}_{(i_1, i_2, t)}^* = 1_{i_1, i_2} \hat{V}_{(i_1, i_2, t)}$ and

$$\begin{aligned} \mathbb{K}_b((x_1, x_2, x_3), (y_1, y_2, y_3)) &= \mathbb{K}((x_1/b_1, x_2/b_2, x_3/b_3), (y_1/b_1, y_2/b_2, y_3/b_3)), \\ \mathbb{K}((x_1, x_2, x_3), (y_1, y_2, y_3)) &= K_1(d_{(x_1, x_2), (y_1, y_2)}) K_2(d_{x_3, y_3}). \end{aligned} \quad (29)$$

Under Assumption F1, $\mathbb{K}_b((x_1, x_2, x_3), (y_1, y_2, y_3))$ and $\mathbb{K}((x_1, x_2, x_3), (y_1, y_2, y_3))$ depend on (x_1, x_2, x_3) and (y_1, y_2, y_3) only through $|x_1 - y_1|$, $|x_2 - y_2|$ and $|x_3 - y_3|$.

Assumption F2 (i) Assumption I7(i) holds. (ii) Either $\mathbb{K}_b(\cdot, \cdot)$ is continuous on $([0, 1]^3 \times [0, 1]^3)$ and continuously differentiable almost everywhere on $([0, 1]^3 \times [0, 1]^3)$ or $\mathbb{K}((x_1, x_2, x_3), (y_1, y_2, y_3)) = 1\{d_{(x_1, x_2), (y_1, y_2)} \leq 1\} \times 1\{d_{x_3, y_3} \leq 1\}$.

Assumption F2 accommodates commonly used kernels. We have to single out the rectangular kernel as $\mathbb{K}_b(\cdot, \cdot)$ is not continuous.

Under Assumption F2(ii), $\mathbb{K}_b(\cdot, \cdot)$ is square integrable on $([0, 1]^3 \times [0, 1]^3)$. So $\mathbb{K}_b(\cdot, \cdot)$ has a Fourier series representation:

$$\mathbb{K}_b((x_1, x_2, x_3), (y_1, y_2, y_3)) = \lim_{N \rightarrow \infty} \sum_{k, \ell, m = -N}^N \tilde{\lambda}_{k, \ell, m} \varphi_{b_1, k}(x_1 - y_1) \varphi_{b_2, \ell}(x_2 - y_2) \varphi_{b_3, m}(x_3 - y_3) \quad (30)$$

where $\varphi_{b_i, k}(r) = \exp(\sqrt{-1} \pi r k / b_i)$ and $\{\tilde{\lambda}_{k, \ell, m}\}$ are the (complex) Fourier coefficients. Using the function form in (29) and Assumption F1, we can rewrite the above complex exponential form as a trigonometric series:

$$\mathbb{K}_b((x_1, x_2, x_3), (y_1, y_2, y_3)) = \lim_{\mathcal{L} \rightarrow \infty} \sum_{\iota=1}^{\mathcal{L}} \lambda_{\iota} \Phi_{b, \iota}((x_1, x_2, x_3)) \Phi_{b, \iota}(y_1, y_2, y_3), \quad (31)$$

where $\{\Phi_{b, \iota}(x_1, x_2, x_3) \Phi_{b, \iota}(y_1, y_2, y_3)\}$ is an orthonormal basis for $L^2([0, 1]^3 \times [0, 1]^3)$ under the Lebesgue measure and by default we set $\Phi_{b, 1}(\cdot)$ to be the constant function.

When $\mathbb{K}_b(\cdot, \cdot)$ is continuous and continuously differentiable almost everywhere on $([0, 1]^3 \times [0, 1]^3)$, the convergence in (30) and (31) is absolute and uniform in $(x_1, x_2, x_3) \in [0, 1]^3$ and $(y_1, y_2, y_3) \in [0, 1]^3$. When $\mathbb{K}_b((x_1, x_2, x_3), (y_1, y_2, y_3))$ is positive definite, continuous and symmetric, the uniform series representation in (31) can also be obtained using Mercer's theorem. When a rectangular kernel is used, the convergence is in terms of the L_1 and L_2 norms under the Lebesgue measure.

Let $V_{(i_1, i_2, t)}^* = 1_{i_1, i_2} V_{(i_1, i_2, t)}$ and $e_{(i_1, i_2, t)}^* = 1_{i_1, i_2} e_{(i_1, i_2, t)}$ with $e_{(i_1, i_2, t)} \stackrel{i.i.d.}{\sim} N(0, I_p)$. We maintain the following high level assumption.

Assumption F3 As $(n, T) \rightarrow \infty$, the following holds

$$\begin{aligned} & P \left(\left[\frac{1}{\sqrt{nT}} \sum_{i_1, i_2, t} \Phi_{b, \iota} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) V_{(i_1, i_2, t)}^* \right] < v \text{ for } \iota = 1, 2, \dots, \mathcal{L} \right) \\ &= P \left(\left[\Lambda \frac{1}{\sqrt{nT}} \sum_{i_1, i_2, t} \Phi_{b, \iota} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) e_{(i_1, i_2, t)}^* \right] < v \text{ for } \iota = 1, 2, \dots, \mathcal{L} \right) + o(1) \end{aligned}$$

for every fixed \mathcal{L} where $v \in \mathbb{R}^p$, $b \in (0, 1]^3$ and Λ is the matrix square root of J , i.e. $\Lambda \Lambda' = J$.

Assumption F3 is satisfied if a CLT holds jointly over $\iota = 1, 2, \dots, \mathcal{L}$ for

$$\frac{1}{\sqrt{nT}} \sum_{i_1, i_2, t} \Phi_{b, \iota} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) V_{(i_1, i_2, t)}^*.$$

This is in contrast with the FCLT assumption often made in the fixed-smoothing asymptotic theory. The above CLT assumption corresponds to the finite dimensional convergence in an FCLT. It is weaker than an FCLT which requires an additional tightness condition. It is not trivial to verify the tightness condition in a spatial setting, as the indexing sets are more complicated than in a time series setting. The CLT holds under weaker conditions and therefore can accommodate a wider range of panel data processes. Some primitive sufficient conditions for this assumption are provided in Sun and Kim (2013).

When Assumption F3 holds, we write

$$\frac{1}{\sqrt{nT}} \sum_{i_1, i_2, t} \Phi_{b, \iota} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) V_{(i_1, i_2, t)}^* \stackrel{a}{\sim} \Lambda \frac{1}{\sqrt{nT}} \sum_{i_1, i_2, t} \Phi_{b, \iota} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) e_{(i_1, i_2, t)}^*$$

jointly over $\iota = 1, 2, \dots, \mathcal{L}$ where ‘ $\stackrel{a}{\sim}$ ’ signifies that the two sides are asymptotically equivalent in distribution as $(n, T) \rightarrow \infty$.

To establish the fixed-smoothing asymptotics, we make one more high level assumption:

Assumption F4 For all $(r_1, r_2, \tau) \in [0, 1]^3$,

$$\frac{1}{nT} \sum_{i_1=1}^{[r_1 L_n]} \sum_{i_2=1}^{[r_2 M_n]} \sum_{t=1}^{[\tau T]} \tilde{Z}_{(i_1, i_2, t)}^* \tilde{X}_{(i_1, i_2, t)}^{*'} \rightarrow^p r_1 r_2 \tau Q$$

where $\tilde{Z}_{(i_1, i_2, t)}^* = 1_{i_1, i_2} \tilde{Z}_{(i_1, i_2, t)}$ and $\tilde{X}_{(i_1, i_2, t)}^* = 1_{i_1, i_2} \tilde{X}_{(i_1, i_2, t)}$.

Proposition 2 Let Assumptions 1 and F1–F4 hold, then for $b \in (0, 1]^3$ we have

$$\hat{J}_{nT} \stackrel{a}{\sim} \Lambda \hat{J}_{nT}^a \Lambda' \quad (32)$$

where for $\bar{e}^* = (nT)^{-1} \sum_{j_1, j_2, s} e_{(j_1, j_2, s)}^*$,

$$\hat{J}_{nT}^a := \frac{1}{nT} \sum_{i_1, j_1, i_2, j_2, t, s} \mathbb{K}_b \left(\left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right), \left(\frac{j_1}{L_n}, \frac{j_2}{M_n}, \frac{s}{T} \right) \right) \left(e_{(i_1, i_2, t)}^* - \bar{e}^* \right) \left(e_{(j_1, j_2, s)}^* - \bar{e}^* \right)'.$$

In the absence of an FCLT, we can not follow the standard arguments for establishing the fixed-smoothing asymptotics to prove Proposition 2. Instead, we rely on Lemma 2 given in the appendix. The lemma is crucial for our proof and may be of independent interest. The demeaning of $e_{(i_1, i_2, t)}^*$ in Proposition 2 reflects the estimation uncertainty in $\hat{\beta}$.

Under Assumptions 1, F3 and F4, we have

$$\sqrt{nT} (\hat{\beta} - \beta_0) \stackrel{a}{\sim} Q^{-1} \Lambda \frac{1}{\sqrt{nT}} \sum_{i_1, i_2, t} e_{(i_1, i_2, t)}^* \quad (33)$$

and this holds jointly with (32). So under H_0 ,

$$\begin{aligned} F_{nT} &\stackrel{a}{\sim} \left(\mathcal{R} Q^{-1} \Lambda \frac{1}{\sqrt{nT}} \sum_{i_1, i_2, t} e_{(i_1, i_2, t)}^* \right)' \left(\mathcal{R} Q^{-1} \Lambda \hat{J}_{nT}^a \Lambda' Q'^{-1} \mathcal{R}' \right)^{-1} \left(\mathcal{R} Q^{-1} \Lambda \frac{1}{\sqrt{nT}} \sum_{i_1, i_2, t} e_{(i_1, i_2, t)}^* \right) / g \\ &\stackrel{d}{=} \eta' \left[\frac{1}{nT} \sum \mathbb{K}_b \left(\left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right), \left(\frac{j_1}{L_n}, \frac{j_2}{M_n}, \frac{s}{T} \right) \right) \left(\varepsilon_{(i_1, i_2, t)}^* - \bar{\varepsilon}^* \right) \left(\varepsilon_{(j_1, j_2, s)}^* - \bar{\varepsilon}^* \right)' \right]^{-1} \eta / g \\ &\stackrel{d}{=} F_{nT}^a(g, b) \end{aligned} \quad (34)$$

where

$$\eta = \frac{1}{\sqrt{nT}} \sum_{i_1, i_2, t} \varepsilon_{(i_1, i_2, t)}^* \text{ and } \varepsilon_{(i_1, i_2, t)}^* = 1_{i_1, i_2} \varepsilon_{(i_1, i_2, t)} \text{ with } \varepsilon_{(i_1, i_2, t)} \stackrel{i.i.d.}{\sim} N(0, I_g).$$

Here we have used $\mathcal{R}Q^{-1}\Lambda e_{(i_1, i_2, t)}^* \stackrel{d}{=} \tilde{\Lambda} \varepsilon_{(i_1, i_2, t)}^*$ for some nonsingular $g \times g$ matrix $\tilde{\Lambda}$. It is easy to show that

$$F_{nT}^a(g, b) = \sqrt{nT} \varepsilon' \left[\frac{1}{nT} \sum_{i,j} \sum_{s,t} K_1 \left(\frac{d_{ij}}{d_n} \right) K_2 \left(\frac{d_{ts}}{d_T} \right) (\varepsilon_{it} - \bar{\varepsilon}) (\varepsilon_{js} - \bar{\varepsilon})' \right]^{-1} \sqrt{nT} \bar{\varepsilon} / g, \quad (35)$$

where $\varepsilon_{it} \stackrel{i.i.d.}{\sim} N(0, I_g)$ and $\bar{\varepsilon} = (nT)^{-1} \sum_{it} \varepsilon_{it}$.

We summarize the above result in the theorem below.

Theorem 5 *Let Assumptions 1 and F1 – F4 hold, then $F_{nT} \stackrel{a}{\sim} F_{nT}^a(g, b)$. More precisely,*

$$P(F_{nT} < x) = P(F_{nT}^a(g, b) < x) + o(1) \text{ as } (n, T) \rightarrow \infty \text{ for fixed } b.$$

Under the fixed-smoothing asymptotics, F_{nT} is asymptotically equivalent in distribution to $F_{nT}^a(g, b)$, which is a quadratic form in a normal vector $\sqrt{nT} \bar{\varepsilon}$ with a random and independent weighting matrix. The random weighting matrix reflects the estimating uncertainty of the variance estimator. The distribution of $F_{nT}^a(g, b)$ is nonstandard but can be easily simulated. To obtain a realization of $F_{nT}^a(g, b)$, we only have to draw nT *i.i.d.* standard normal g -vectors $\{\varepsilon_{it}\}$ and plug them into the simple representation in (35). Lattice mapping, which is needed for our theoretical development, is not necessary in empirical implementation of our test.

7.2 Expansion of $F_{nT}^a(g, b)$ and F -approximation

Under the sequential asymptotics where $(n, T) \rightarrow \infty$ for fixed b_1, b_2, b_3 followed by letting $(b_1, b_2, b_3) \rightarrow 0$, we present the asymptotic expansion of the distribution of $F_{nT}^a(g, b)$ in (34) and establish the validity of a standard F -approximation.

Define the centered version of the kernel function $\mathbb{K}_b^*(\cdot, \cdot)$ as

$$\begin{aligned} & \mathbb{K}_b^*((x_1, x_2, x_3), (y_1, y_2, y_3)) \\ &= \mathbb{K}_b((x_1, x_2, x_3), (y_1, y_2, y_3)) - \int_{[0,1]^3} \mathbb{K}_b((x_1, x_2, x_3), (y_1, y_2, y_3)) dx_1 dx_2 dx_3 \\ & \quad - \int_{[0,1]^3} \mathbb{K}_b((x_1, x_2, x_3), (y_1, y_2, y_3)) dy_1 dy_2 dy_3 \\ & \quad + \int_{[0,1]^3} \int_{[0,1]^3} \mathbb{K}_b((x_1, x_2, x_3), (y_1, y_2, y_3)) dx_1 dx_2 dx_3 dy_1 dy_2 dy_3. \end{aligned}$$

Then it is easy to show that

$$\hat{J}_{nT} \stackrel{a}{\sim} \Lambda \left[\frac{1}{nT} \sum_{i_1, j_1, i_2, j_2, t, s} \mathbb{K}_b^* \left(\left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right), \left(\frac{j_1}{L_n}, \frac{j_2}{M_n}, \frac{s}{T} \right) \right) e_{(i_1, i_2, t)}^* e_{(j_1, j_2, s)}^{*'} \right]^{-1} \Lambda'. \quad (36)$$

Since $\mathbb{K}_b^*((x_1, x_2, x_3), (y_1, y_2, y_3)) \in L^2([0, 1]^6)$, it has a Fourier series representation:

$$\begin{aligned} & \mathbb{K}_b^*((x_1, x_2, x_3), (y_1, y_2, y_3)) \\ &= \sum_{k, \ell, m, k', \ell', m'=1}^{\infty} \lambda_{k\ell m k' \ell' m'} \psi_{b_1, k}(x_1) \psi_{b_2, \ell}(x_2) \psi_{b_3, m}(x_3) \psi_{b_1, k'}(y_1) \psi_{b_2, \ell'}(y_2) \psi_{b_3, m'}(y_3) \\ &:= \sum_{k, \ell, m, k', \ell', m'=1}^{\infty} \lambda_{k\ell m k' \ell' m'} \varrho_{b, k\ell m}(x_1, x_2, x_3) \varrho_{b, k' \ell' m'}(y_1, y_2, y_3), \end{aligned} \quad (37)$$

where $\{\varrho_{b,k\ell m}(x_1, x_2, x_3) \varrho_{b,k'\ell'm'}(y_1, y_2, y_3)\}$ is an orthonormal basis for $L^2([0, 1]^3 \times [0, 1]^3)$. The convergence is in terms of L_1 and L_2 norms, which is sufficient for a distributional representation. As $\int_{[0,1]^3} \mathbb{K}_b^*((x_1, x_2, x_3), (y_1, y_2, y_3)) dx_1 dx_2 dx_3 = 0$ for any (y_1, y_2, y_3) by definition, $\varrho_{b,k\ell m}(\cdot)$ has the ‘zero mean’ property, i.e.

$$\int_{[0,1]^3} \varrho_{b,k\ell m}(x_1, x_2, x_3) dx_1 dx_2 dx_3 = 0.$$

Using the representation in (37), we have

$$\begin{aligned} & \frac{1}{nT} \sum_{i_1, j_1, i_2, j_2, t, s} \mathbb{K}_b^* \left(\left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right), \left(\frac{j_1}{L_n}, \frac{j_2}{M_n}, \frac{s}{T} \right) \right) \varepsilon_{(i_1, i_2, t)}^* \varepsilon_{(j_1, j_2, s)}^{*'} \\ & \stackrel{d}{=} \sum_{k, \ell, m, k', \ell', m'=1}^{\infty} \lambda_{k\ell m k' \ell' m'} \xi_{b, k\ell m} \xi_{b, k' \ell' m'}', \end{aligned} \quad (38)$$

where $\xi_{b, k\ell m} = (nT)^{-1/2} \sum_{i_1, i_2, t} \varrho_{b, k\ell m}(i_1/L_n, i_2/M_n, t/T) \varepsilon_{(i_1, i_2, t)}^*$.

We can simplify the above representation. First, using the Cantor tuple function, we can encode (h_1, h_2, h_3) into a single natural number h . That is,

$$h = \pi^{(3)}(h_1, h_2, h_3) := \pi^{(2)}(\pi^{(2)}(h_1, h_2), h_3),$$

where

$$\pi^{(2)}(h_1, h_2) = \frac{1}{2}(h_1 + h_2)(h_1 + h_2 + 1) + h_2.$$

The map between (h_1, h_2, h_3) and h is one-to-one and onto. With this definition, we abuse the notation a little and write

$$\lambda_{h_1 h_2 h_3 h'_1 h'_2 h'_3} = \lambda_{hh'} \text{ and } \xi_{b, h_1 h_2 h_3} = \xi_{b, h}.$$

With this result, we follow Sun and Kim (2012) to obtain

$$\sum_{k, \ell, m, k', \ell', m'=1}^{\infty} \lambda_{k\ell m k' \ell' m'} \xi_{b, k\ell m} \xi_{b, k' \ell' m'}' = \sum_{k=1}^{\infty} \lambda_k^* \zeta_{nT, k} \zeta_{nT, k}'',$$

where λ_k^* is related to the centered kernel function $\mathbb{K}_b^*(\cdot, \cdot)$ and $\zeta_{nT, k} \xrightarrow{d} \zeta_k \stackrel{i.i.d.}{\sim} N(0, I_g)$.

Using Lemma 2 in the appendix, we can show that for fixed b_1, b_2, b_3 ,

$$gF_{nT}^a(g, b) \xrightarrow{d} gF_{\infty}(g, b) \stackrel{d}{=} \phi' \left[\sum_{k=1}^{\infty} \lambda_k^* \zeta_k \zeta_k' \right]^{-1} \phi, \text{ as } (n, T) \rightarrow \infty$$

where $\phi \sim N(0, I_g)$ and ϕ is independent of ζ_k for all k . By definition, $\zeta_k \zeta_k'$ is a Wishart distribution $\mathbb{W}_g(I_g, 1)$, so $\sum_{k=1}^{\infty} \lambda_k^* \zeta_k \zeta_k'$ is an infinite weighted sum of independent Wishart distributions.

Let

$$\sum_{k=1}^{\infty} \lambda_k^* \zeta_k \zeta_k' = \begin{pmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{pmatrix},$$

where v_{11} is a scalar. Following Sun (2010), we can show that

$$P\{gF_{\infty}(g, b) \leq z\} = EG_g(z(v_{11} - v_{12}v_{22}^{-1}v_{21})) = EG_g(zv_{11.2}),$$

where $G_g(\cdot)$ is the cdf of a central χ_g^2 variate and $v_{11.2} = v_{11} - v_{12}v_{22}^{-1}v_{21}$. As $(b_1, b_2, b_3) \rightarrow 0$, we expect $v_{11.2}$ to be concentrated around 1. By taking a Taylor expansion $G_g(zv_{11.2})$ around $G_g(z)$ and computing the moments of $v_{11.2}$, we can prove the following theorem.

Theorem 6 *Suppose Assumptions 1 and F1–F4 hold. Under the sequential asymptotics where $(n, T) \rightarrow \infty$ followed by letting $(b_1, b_2, b_3) \rightarrow 0$, we have*

$$\lim_{(n, T) \rightarrow \infty} P\{gF_{nT}^a(g, b) \leq z\} = P\{gF_\infty(g, b) \leq z\} = G_g(z) + A(z)b_1b_2b_3 + o(b_1b_2b_3)$$

where

$$A(z) = G_g''(z)z^2c_2 - G_g'(z)z[c_1 + (g-1)c_2],$$

and

$$\begin{aligned} c_1 &= \int_{[-1,1]^3} \mathbb{K}((x_1, x_2, x_3), (x_1 + y_1, x_2 + y_2, x_3 + y_3)) dy_1 dy_2 dy_3, \\ c_2 &= \int_{[-1,1]^3} \mathbb{K}^2((x_1, x_2, x_3), (x_1 + y_1, x_2 + y_2, x_3 + y_3)) dy_1 dy_2 dy_3. \end{aligned}$$

Since $\mathbb{K}((x_1, x_2, x_3), (x_1 + y_1, x_2 + y_2, x_3 + y_3))$ depends only on (y_1, y_2, y_3) , c_1 and c_2 are constants, which can be computed either analytically or numerically.

Theorem 6 characterizes the nonstandard distribution $gF_\infty(g, b)$ when b_1, b_2 and b_3 are small. It clearly shows that the difference between $gF_\infty(g, b)$ and χ_g^2 depends on the smoothing parameters, kernel function and the number of restrictions being tested.

It is interesting to see that this representation of $gF_\infty(g, b)$ is the same as that obtained by Sun (2010) for the fixed-smoothing asymptotic distribution of the Wald statistic in a time series context.

Let

$$\begin{aligned} \mu_1 &= \sum_{k=1}^{\infty} \lambda_k^* = \int_{[0,1]^3} \mathbb{K}_b^*((x_1, x_2, x_3), (x_1, x_2, x_3)) dx_1 dx_2 dx_3 \\ &= 1 - \int_{[0,1]^3} \int_{[0,1]^3} \mathbb{K}_b((x_1, x_2, x_3), (y_1, y_2, y_3)) dx_1 dx_2 dx_3 dy_1 dy_2 dy_3 \\ \mu_2 &= \sum_{k=1}^{\infty} (\lambda_k^*)^2 = \int_{[0,1]^3} \int_{[0,1]^3} [\mathbb{K}_b^*((x_1, x_2, x_3), (y_1, y_2, y_3))]^2 dx_1 dx_2 dx_3 dy_1 dy_2 dy_3. \end{aligned}$$

Define $D = \lceil \mu_1^2 / \mu_2 \rceil$ where $\lceil \cdot \rceil$ denotes the ceiling function. Then using the same argument as in Sun (2010), we have the following approximation:

$$\frac{\mu_1(D - g + 1)}{D} F_\infty(g, b) \stackrel{d}{\approx} F_{g, D-g+1}. \quad (39)$$

The following theorem gives a rigorous description of the F -approximation.

Theorem 7 *Suppose Assumptions 1 and F1–F4 hold. As $(b_1, b_2, b_3) \rightarrow 0$, we have*

$$P\left\{\frac{\mu_1(D - g + 1)}{D} F_\infty(g, b) \leq z\right\} = P\{F_{g, D-g+1} \leq z\} + o(b_1b_2b_3).$$

To use Theorem 7, we can estimate D by $D_{nT} = \lceil \mu_{1nT}^2 / \mu_{2nT} \rceil$ where

$$\mu_{1nT} = 1 - \frac{1}{(nT)^2} \sum_{i,j,t,s} K_1 \left(\frac{d_{ij}}{d_n} \right) K_2 \left(\frac{d_{ts}}{d_T} \right), \quad \mu_{2nT} = \frac{1}{(nT)^2} \sum_{i,j,t,s} (K_{i,j,t,s}^*)^2, \quad (40)$$

and

$$\begin{aligned} K_{i,j,t,s}^* &= K_1 \left(\frac{d_{ij}}{d_n} \right) K_2 \left(\frac{d_{ts}}{d_T} \right) - \frac{1}{nT} \sum_{\tilde{j}, \tilde{s}} K_1 \left(\frac{d_{i\tilde{j}}}{d_n} \right) K_2 \left(\frac{d_{t\tilde{s}}}{d_T} \right) \\ &\quad - \frac{1}{nT} \sum_{\tilde{i}, \tilde{t}} K_1 \left(\frac{d_{\tilde{i}j}}{d_n} \right) K_2 \left(\frac{d_{\tilde{t}s}}{d_T} \right) + \frac{1}{(nT)^2} \sum_{\tilde{i}, \tilde{t}} \sum_{\tilde{j}, \tilde{s}} K_1 \left(\frac{d_{\tilde{i}\tilde{j}}}{d_n} \right) K_2 \left(\frac{d_{\tilde{t}\tilde{s}}}{d_T} \right). \end{aligned}$$

Here μ_{1nT} , μ_{2nT} and $K_{i,j,t,s}^*$ are the finite sample versions of μ_1 , μ_2 and K_b^* , respectively.

It is easy to show that $\mu_{1nT} = \mu_1 (1 + o(1))$, $\mu_{2nT} = \mu_2 (1 + o(1))$ and $D_{nT} = D (1 + o(1))$, as $(n, T) \rightarrow \infty$. Let $D_{nT}^* = \max(5, \lceil D_{nT} - g + 1 \rceil)$ and $F_{g, D_{nT}^*}^\alpha$ be the $1 - \alpha$ quantile of the F distribution with the degrees of freedom g and D_{nT}^* . Based on Theorem 7, for the F -test version of Wald statistic, F_{nT} , we can use

$$\mathcal{F}_{nT}^\alpha := \nu_{nT} F_{g, D_{nT}^*}^\alpha \quad \text{where} \quad \nu_{nT} = \frac{D_{nT}}{\mu_{1nT} \max\{1, (D_{nT} - g + 1)\}}, \quad (41)$$

as the critical value for the test with nominal size α . In (41), we employ D_{nT}^* and $\max\{1, (D_{nT} - g + 1)\}$ in place of $D_{nT} - g + 1$ to ensure that the variance of the F distribution exists and that ν_{nT} is positive. We use the critical values in (41) in our simulation.

Unreported simulation results indicate that \mathcal{F}_{nT}^α are reasonably close to the $1 - \alpha$ quantile of $F_{nT}^a(g, b)$ when b_1, b_2 and b_3 are small (< 0.3). Accordingly, we recommend using the adjusted F critical values \mathcal{F}_{nT}^α when the data-driven bandwidths turn out to be small. As b_1, b_2 and b_3 increase, however, the discrepancy of the F -approximation from $F_{nT}^a(g, b)$ may become large. Thus, if the bandwidth selection rule yields large bandwidths, we recommend using the nonstandard critical values obtained by simulating the asymptotically equivalent distribution given in (35).

8 Monte Carlo simulation

In this section, we provide some simulation evidence on the finite sample performance of our covariance estimator and the associated testing procedure. We choose the bandwidths based on the AMSE* criterion and consider the rectangular kernel as well as the Parzen kernel to construct \hat{J}_{nT} . We compare the performance of \hat{J}_{nT} with \hat{J}_{nT}^{DK} , \hat{J}_{nT}^A and \hat{J}_{nT}^{DK*} . We evaluate the covariance estimators and the associated testing procedures using the RMSE criterion, the coverage error of the associated confidence intervals (CIs) or regions, and the size-adjusted power. The coverage error of the CIs is equivalent to the error of rejection probability of the underlying tests under the null. We examine the robustness to the measurement errors in economic distance. It is also investigated how the number of restrictions being tested affects the performance of the Wald test under the two different limiting thought experiments.

We assume a lattice structure, in which each individual is located on a square grid of integers. We use the Euclidean distance for d_{ij} . The data generating processes we consider here are:

$$\begin{aligned}
\text{DGP1: } & Y_{it} = \beta_0 + u_{it} \quad \beta_0 = 0; \\
& u_t = \lambda u_{t-1} + \varepsilon_t, \quad \varepsilon_t = (I - \theta \tilde{W}_n)^{-1} v_t, \quad v_t \stackrel{i.i.d.}{\sim} N(0, I_n); \\
\text{DGP2: } & Y_{it} = X_{it}^{(1)} \beta_{10} + \dots + X_{it}^{(p)} \beta_{p0} + \alpha_i + f_t + u_{it}, \\
& \beta_{10} = \dots = \beta_{p0} = 0, \quad \alpha_i = f_t = 0; \\
& X_t = \lambda X_{t-1} + \nu_t, \quad \nu_t = (I - \theta \tilde{W}_n)^{-1} \eta_t, \quad \eta_t \stackrel{i.i.d.}{\sim} N(0, I_n) \\
& u_t = \lambda u_{t-1} + \varepsilon_t, \quad \varepsilon_t = (I - \theta \tilde{W}_n)^{-1} v_t, \quad v_t \stackrel{i.i.d.}{\sim} N(0, I_n),
\end{aligned}$$

where X_{it} is a p -vector, $X_t = (X_{1t}, \dots, X_{nt})'$ and $u_t = (u_{1t}, \dots, u_{nt})'$. \tilde{W}_n is a contiguity matrix and individuals i and j are neighbors if $d_{ij} = 1$. Following the convention, it is row-standardized and its diagonal elements are zero. The parameters λ and θ determine the strength of the temporal and spatial correlation. We consider the following values for λ and θ : 0, 0.3, 0.6 and 0.9.

DGP1 is used for the RMSE criterion and DGP2 is used for the coverage accuracy of the associated CIs. DGP2 includes the individual and time effects and β_0 is estimated with the fixed-effects OLS estimator. It is a special case of the setting of this paper in (1) in which $X_{it} = Z_{it}$. In contrast, these effects are absent in DGP1 for easy calculation of the RMSE. We estimate β_0 in DGP1 by the sample average.

For the estimators \hat{J}_{nT}^{DK} and \hat{J}_{nT}^{DK*} , we employ the respective data-driven bandwidths in (27) and (28), using the time series AR(1) or spatial AR(1) as the approximating plug-in model. For \hat{J}_{nT} with the Parzen kernel, we employ the bandwidths given in (14) and (15), using the spatiotemporal parametric model in (20) as the approximating plug-in model. W_n is the contiguity matrix in which individuals i and j are neighbors if $d_{ij} = 1$. We set $\eta_n = 2$ and $\ell_n = \pi d_n^2$. Note that the approximating parametric models for \hat{J}_{nT}^{DK*} and \hat{J}_{nT} are mis-specified whereas the AR(1) model for \hat{J}_{nT}^{DK} is correctly specified. We employ the QMLE to estimate parameters in (20) and (28). For \hat{J}_{nT} with the rectangular kernel, we use the Parzen kernel as the target kernel to obtain the data-driven bandwidths.

To obtain a positive semi-definite covariance estimator with the rectangular kernel, we follow Politis (2011) and modify \hat{J}_{nT} . According to the spectral decomposition, $\hat{J}_{nT} = \hat{U} \hat{\Lambda} \hat{U}'$, where \hat{U} is an orthogonal matrix and $\hat{\Lambda} = \text{diag}(\hat{\lambda}_1, \dots, \hat{\lambda}_p)$ is a diagonal matrix whose diagonal elements are the eigenvalues of \hat{J}_{nT} . Let $\hat{\Lambda}^+ = \text{diag}(\hat{\lambda}_1^+, \dots, \hat{\lambda}_p^+)$ where $\hat{\lambda}_s^+ = \max(\hat{\lambda}_s, 0)$. Then, we define our modified estimator as $\hat{J}_{nT}^+ = \hat{U} \hat{\Lambda}^+ \hat{U}'$. As each eigenvalue of \hat{J}_{nT}^+ is nonnegative, \hat{J}_{nT}^+ is positive semi-definite.

The number of simulation replications is 5000, and three different sample sizes are considered; (i) small T and n ; $T = 15, n = 49$ (7×7), (ii) large T and small n ; $T = 50, n = 49$, and (iii) small T and large n ; $T = 15, n = 196$ (14×14). The following values are used for each kernel.

	\bar{K}_1	\bar{K}_2	c_1	c_2	K_q
Parzen	0.2889	0.2697	0.4123	0.1558	-6
Rectangular	1	1	6.2926	6.2926	

We allow for the case with measurement errors in the distance measure. The error contaminated distance, d_{ij}^* , is generated as follows. If $d_{ij} < 2$, then d_{ij} is observed without a measurement error. If $d_{ij} \geq 2$, then we observe d_{ij}^* :

$$d_{ij}^* = d_{ij} + e_{ij},$$

where $e_{ij} = -1, 0, 1$ with equal probabilities. PHAC, CCE, DK and DK* denote the test statistics based on \hat{J}_{nT} , \hat{J}_{nT}^A , \hat{J}_{nT}^{DK} , and \hat{J}_{nT}^{DK*} , respectively. We use the F -approximation based on (41) to obtain critical values under the fixed-smoothing asymptotics.

Table 1 presents the ratios of the RMSE to J_{nT} for \hat{J}_{nT} and \hat{J}_{nT}^{DK} evaluated at the data dependent bandwidth parameters (\hat{d}_n, \hat{d}_T) and \hat{d}_T^{DK} and at infeasible optimal bandwidth parameters (d_n^*, d_T^*) and d_T^{DK} . The infeasible bandwidth parameters are obtained by plugging the true data generating process into the AMSE* and AMSE formulae. Several patterns emerge. First, \hat{J}_{nT} outperforms \hat{J}_{nT}^{DK} in almost all the cases. When spatial dependence is absent or weak, \hat{J}_{nT} has a substantially smaller RMSE than \hat{J}_{nT}^{DK} . Even when $\theta = 0.9$, these two estimators are not much different. In particular, when the rectangular kernel is used, \hat{J}_{nT} is as accurate as and sometimes more accurate than \hat{J}_{nT}^{DK} . This implies that adaptiveness works well in this setting. Second, increasing n reduces only the RMSE of \hat{J}_{nT} while increasing T reduces the RMSEs of both estimators. This is expected, as the rate of convergence of \hat{J}_{nT}^{DK} depends only on T while that of \hat{J}_{nT} depends on both n and T . Finally, the results under both feasible and infeasible AMSE*-optimal bandwidths show that the AMSE* criterion is effective in controlling the RMSE of \hat{J}_{nT} .¹⁰

Table 2 reports the empirical coverage probabilities (ECPs) of 95% CIs associated with the different covariance estimators: \hat{J}_{nT} , \hat{J}_{nT}^{DK} , \hat{J}_{nT}^A , and \hat{J}_{nT}^{DK*} . DGP2 is used with a univariate regressor ($p = 1$). We use both the fixed-smoothing asymptotics and the increasing-smoothing asymptotics for critical values based on PHAC. We use the fixed-smoothing asymptotics for DK and DK*, and the increasing-smoothing asymptotics for CCE. From this table, we can first compare the size properties of PHAC using the rectangular kernel with those of the alternative test statistics. When $\theta = 0$ with high temporal autocorrelation, CCE performs better than PHAC. However, as θ increases, the performance of PHAC becomes better than that of CCE. Compared with DK*, the CIs associated with PHAC have more accurate coverage probability if the process is spatially persistent. When the process is temporally persistent, DK* yields more accurate coverage probability. Both PHAC and DK* become more accurate with large n , but only the performance of PHAC improves when T increases. In comparison with DK, we see that PHAC is more accurate when the process is temporally persistent or n is large. When a process is spatially persistent and temporal dependence is weak, DK tends to show better performance in testing, but PHAC also performs almost as good as DK. Second, Table 2 compares the performances of PHAC under two different asymptotics. The results indicate that the fixed-smoothing asymptotic approximation is substantially more accurate than the increasing-smoothing asymptotic approximation. The difference increases as the process becomes more persistent. When $\theta = \lambda = 0.9$ and $T = 15, n = 49$, the ECP of the PHAC with the rectangular kernel under the fixed-smoothing asymptotics is 80.0% but it is only 63.0% under the increasing-smoothing asymptotics. Third, Table 2 provides strong evidence that the rectangular kernel performs better than the finite-order kernel under the fixed-smoothing asymptotics. The performance of PHAC with the rectangular kernel is very robust to spatial dependence so that the size distortion does not increase much with spatial dependence. This size advantage of the rectangular kernel arises from its bias reducing property and the adaptiveness of the bandwidth choice rule. Finally, Table 2 shows that our testing procedure based on the fixed-smoothing asymptotics is reasonably robust to measurement errors. Comparing PHAC with PHAC_e, we see that the performance of PHAC_e is quite close to that of PHAC in most cases.

Table 3 compares the performances of the two different asymptotics when more than one

¹⁰The RMSE of \hat{J}_{nT}^{DK*} has also been compared in the simulation. Unreported results show that \hat{J}_{nT} tends to have a smaller RMSE than \hat{J}_{nT}^{DK*} in most cases and especially with large T and/or under weak temporal autocorrelation.

parameters or restrictions are considered. DGP2 is used with $p = 3$. The confidence regions are obtained by inverting the Wald test of $H_0 : \beta_1 = 0$ with $g = 1$ and $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ with $g = 3$, respectively. The table evidently indicates that under the increasing-smoothing asymptotics the error in coverage probability increases with the number of parameters being considered. The coverage error becomes especially severe when the process is highly persistent. When $g = 3$ and $\theta = \lambda = 0.9$, the ECP of PHAC with the Parzen kernel is only 28.5% under the increasing-smoothing asymptotics. The coverage error of PHAC also increases under the fixed-smoothing asymptotics with the number of parameters or restrictions being tested but much lesser. This is consistent with our asymptotic expansion in Theorem 6. The theorem shows that the fixed-smoothing asymptotics and F -approximation correct for the number of restrictions being jointly tested.

Figure 3 presents size-adjusted power of the PHAC and DK with the sample size of $T = 15, n = 49$. We use the DGP2 with $p = 1$, but consider the following local alternative hypothesis

$$H_1(\delta^2) : \beta_1 = \beta_{10} + c/\sqrt{nT}$$

where $c = \left(E \left(\tilde{X}_{it} \tilde{X}_{it}' \right)^{-1} \left[(nT)^{-1} \sum_{i,t} \sum_{j,s} E \left(u_{it} u_{js} \tilde{X}_{it} \tilde{X}_{js}' \right) \right] E \left(\tilde{X}_{it} \tilde{X}_{it}' \right)^{-1} \right)^{-1/2} \tilde{c}$ with $\tilde{c} = \delta\psi / \|\psi\|, \psi \stackrel{i.i.d.}{\sim} N(0, 1)$. The scaling matrix c is computed by simulation. We compute the power using the 5% empirical critical values under the null and with data-driven bandwidth parameters. Figure 3 show that the proposed procedure has better power in most of dependence structures we consider. Even under strong spatial dependence ($\theta = 0.9$), it has almost the same power as the DK except one extreme case ($\theta = 0.9, \lambda = 0.9$).

9 Conclusion

In this paper we study robust inference for linear panel models with fixed effects in the presence of heteroskedasticity and spatiotemporal dependence of unknown forms. We consider a bivariate kernel covariance matrix estimator and examine the properties of the covariance estimator and the associated test statistic under both the increasing-smoothing asymptotics and the fixed-smoothing asymptotics. We also derive the optimal bandwidth selection procedure based on an upper bound of the AMSE. For the fixed-smoothing asymptotic distribution, we establish the validity of an F -approximation. The adaptiveness of our estimator ensures that it can be safely used without the knowledge of the dependence structure.

Instead of using the upper bound of the AMSE as the criterion, we can study the optimal bandwidth selection based on a criterion that is most suitable for hypothesis testing and CI construction. It is interesting to extend the bandwidth selection methods in time series HAC estimation by Sun (2010) and Sun and Kaplan (2011) to the panel setting.

APPENDIX

Proof of Theorem 1

For notational simplicity, we re-order the individuals and time and make new indices. For $i_{(j)} = 1, \dots, \ell_{j,n}$, $d_{i_{(j)}j} \leq d_n$, and for $i_{(j)} = \ell_{j+1,n}, \dots, n$, $d_{i_{(j)}j} > d_n$. For $t_{(s)} = 1, \dots, \ell_{s,T}$, $d_{t_{(s)}s} \leq d_T$, and for $t_{(s)} = \ell_{s+1,T}, \dots, T$, $d_{t_{(s)}s} > d_T$.

(a) Asymptotic Variance

We have

$$\frac{nT}{\ell_n \ell_T} \text{cov} \left(\tilde{J}_{nT}(c_1, d_1), \tilde{J}_{nT}(c_2, d_2) \right) := \frac{1}{nT \ell_n \ell_T} (C_{1nT} + C_{2nT} + C_{3nT}),$$

where

$$\begin{aligned} C_{1nT} &= \sum_{l=1}^{nTp} (E\varepsilon_l^4 - 3) \sum_{i,j,a,b=1}^n \sum_{t,s,u,v=1}^T K_1 \left(\frac{d_{ij}}{d_n} \right) K_2 \left(\frac{d_{ts}}{d_T} \right) K_1 \left(\frac{d_{ab}}{d_n} \right) K_2 \left(\frac{d_{uv}}{d_T} \right) r_{(i,t),l}^{(c_1)} r_{(j,s),l}^{(d_1)} r_{(a,u),l}^{(c_2)} r_{(b,v),l}^{(d_2)} \\ C_{2nT} &= \sum_{l,k=1}^{nTp} \sum_{i,j,a,b=1}^n \sum_{t,s,u,v=1}^T K_1 \left(\frac{d_{ij}}{d_n} \right) K_2 \left(\frac{d_{ts}}{d_T} \right) K_1 \left(\frac{d_{ab}}{d_n} \right) K_2 \left(\frac{d_{uv}}{d_T} \right) r_{(i,t),l}^{(c_1)} r_{(j,s),k}^{(d_1)} r_{(a,u),l}^{(c_2)} r_{(b,v),k}^{(d_2)} \\ C_{3nT} &= \sum_{l,k=1}^{nTp} \sum_{i,j,a,b=1}^n \sum_{t,s,u,v=1}^T K_1 \left(\frac{d_{ij}}{d_n} \right) K_2 \left(\frac{d_{ts}}{d_T} \right) K_1 \left(\frac{d_{ab}}{d_n} \right) K_2 \left(\frac{d_{uv}}{d_T} \right) r_{(i,t),l}^{(c_1)} r_{(j,s),k}^{(d_1)} r_{(a,u),k}^{(c_2)} r_{(b,v),l}^{(d_2)} \end{aligned}$$

For C_{1nT} , under Assumptions I1 and I2

$$\frac{1}{nT \ell_n \ell_T} |C_{1nT}| \leq \frac{c_R^4}{\ell_n \ell_T} \frac{1}{nT} \sum_{l=1}^{nTp} |E\varepsilon_l^4 - 3| \leq \frac{c_R^4 c_{EP}}{\ell_n \ell_T} = o(1). \quad (\text{A.1})$$

In order to consider boundary effects, we can decompose C_{2nT} as follows

$$C_{2nT} := D_{1nT} + D_{2nT} + D_{3nT} + D_{4nT} + D_{5nT}$$

where

$$\begin{aligned}
D_{1nT} &= \sum_{i,a \in E_n} \sum_{j(i)=1}^{\ell_{i,n}} \sum_{b(a)=1}^{\ell_{a,n}} \sum_{t,u \in E_T} \sum_{s(t)=1}^{\ell_{t,T}} \sum_{v(u)=1}^{\ell_{u,T}} K_1 \left(\frac{d_{ij(i)}}{d_n} \right) K_1 \left(\frac{d_{ab(a)}}{d_n} \right) K_2 \left(\frac{d_{ts(t)}}{d_T} \right) K_2 \left(\frac{d_{uv(u)}}{d_T} \right) \\
&\quad \times \gamma_{(it,au)}^{(c_1 c_2)} \gamma_{(j(i)s(t),b(a)v(u))}^{(d_1 d_2)} \\
D_{2nT} &= \sum_{i,a=1}^n \sum_{j(i)=1}^{\ell_{i,n}} \sum_{b(a)=1}^{\ell_{a,n}} \sum_{t \notin E_T} \sum_{u=1}^T \sum_{s(t)=1}^{\ell_{t,T}} \sum_{v(u)=1}^{\ell_{u,T}} K_1 \left(\frac{d_{ij(i)}}{d_n} \right) K_1 \left(\frac{d_{ab(a)}}{d_n} \right) K_2 \left(\frac{d_{ts(t)}}{d_T} \right) K_2 \left(\frac{d_{uv(u)}}{d_T} \right) \\
&\quad \times \gamma_{(it,au)}^{(c_1 c_2)} \gamma_{(j(i)s(t),b(a)v(u))}^{(d_1 d_2)} \\
D_{3nT} &= \sum_{i,a=1}^n \sum_{j(i)=1}^{\ell_{i,n}} \sum_{b(a)=1}^{\ell_{a,n}} \sum_{t \in E_T} \sum_{u \notin E_T} \sum_{s(t)=1}^{\ell_{t,T}} \sum_{v(u)=1}^{\ell_{u,T}} K_1 \left(\frac{d_{ij(i)}}{d_n} \right) K_1 \left(\frac{d_{ab(a)}}{d_n} \right) K_2 \left(\frac{d_{ts(t)}}{d_T} \right) K_2 \left(\frac{d_{uv(u)}}{d_T} \right) \\
&\quad \times \gamma_{(it,au)}^{(c_1 c_2)} \gamma_{(j(i)s(t),b(a)v(u))}^{(d_1 d_2)} \\
D_{4nT} &= \sum_{i \notin E_n} \sum_{a=1}^n \sum_{j(i)=1}^{\ell_{i,n}} \sum_{b(a)=1}^{\ell_{a,n}} \sum_{t,u \in E_T} \sum_{s(t)=1}^{\ell_{t,T}} \sum_{v(u)=1}^{\ell_{u,T}} K_1 \left(\frac{d_{ij(i)}}{d_n} \right) K_1 \left(\frac{d_{ab(a)}}{d_n} \right) K_2 \left(\frac{d_{ts(t)}}{d_T} \right) K_2 \left(\frac{d_{uv(u)}}{d_T} \right) \\
&\quad \times \gamma_{(it,au)}^{(c_1 c_2)} \gamma_{(j(i)s(t),b(a)v(u))}^{(d_1 d_2)} \\
D_{5nT} &= \sum_{i \in E_n} \sum_{a \notin E_n} \sum_{j(i)=1}^{\ell_{i,n}} \sum_{b(a)=1}^{\ell_{a,n}} \sum_{t,u \in E_T} \sum_{s(t)=1}^{\ell_{t,T}} \sum_{v(u)=1}^{\ell_{u,T}} K_1 \left(\frac{d_{ij(i)}}{d_n} \right) K_1 \left(\frac{d_{ab(a)}}{d_n} \right) K_2 \left(\frac{d_{ts(t)}}{d_T} \right) K_2 \left(\frac{d_{uv(u)}}{d_T} \right) \\
&\quad \times \gamma_{(it,au)}^{(c_1 c_2)} \gamma_{(j(i)s(t),b(a)v(u))}^{(d_1 d_2)}
\end{aligned}$$

D_{1nT} is based on nonboundary units whereas D_{2nT} , D_{3nT} , D_{4nT} and D_{5nT} are based on boundary ones.

First, applying the proof of Theorem 1 in Kim and Sun (2011), we can show that

$$\begin{aligned}
&\lim_{(n,T) \rightarrow \infty} \frac{1}{nT\ell_n\ell_T} \sum_{i,a \in E_n} \sum_{j(i)=1}^{\ell_{i,n}} \sum_{b(a)=1}^{\ell_{a,n}} \sum_{t,u \in E_T} \sum_{s(t)=1}^{\ell_{t,T}} \sum_{v(u)=1}^{\ell_{u,T}} K_1^2 \left(\frac{d_{ij(i)}}{d_n} \right) K_2^2 \left(\frac{d_{ts(t)}}{d_T} \right) \gamma_{(it,au)}^{(c_1 c_2)} \gamma_{(j(i)s(t),b(a)v(u))}^{(d_1 d_2)} \\
&= \bar{\mathcal{K}}_1 \bar{\mathcal{K}}_2 J(c_1, c_2) J(d_1, d_2), \tag{A.2}
\end{aligned}$$

and

$$\begin{aligned}
&\lim_{(n,T) \rightarrow \infty} \frac{1}{nT\ell_n\ell_T} D_{1nT} \\
&= \lim_{(n,T) \rightarrow \infty} \frac{1}{nT\ell_n\ell_T} \sum_{i,a \in E_n} \sum_{j(i)=1}^{\ell_{i,n}} \sum_{b(a)=1}^{\ell_{a,n}} \sum_{t,u \in E_T} \sum_{s(t)=1}^{\ell_{t,T}} \sum_{v(u)=1}^{\ell_{u,T}} K_1^2 \left(\frac{d_{ij(i)}}{d_n} \right) K_2^2 \left(\frac{d_{ts(t)}}{d_T} \right) \gamma_{(it,au)}^{(c_1 c_2)} \gamma_{(j(i)s(t),b(a)v(u))}^{(d_1 d_2)}. \tag{A.3}
\end{aligned}$$

It is straightforward to show that (A.2) and (A.3) imply

$$\lim_{(n,T) \rightarrow \infty} \frac{1}{nT\ell_n\ell_T} D_{1nT} = \bar{\mathcal{K}}_1 \bar{\mathcal{K}}_2 J(c_1, c_2) J(d_1, d_2).$$

For D_{2nT} , we have

$$\begin{aligned} & \frac{1}{nT\ell_n\ell_T} D_{2nT} \\ & \leq \frac{1}{nT} \sum_{i,a=1}^n \sum_{t \notin E_T} \sum_{u=1}^T \left| \gamma_{(it,au)}^{(c_1c_2)} \right| \left(\frac{1}{\ell_n\ell_T} \sum_{j_{(i)}=1}^{\ell_{i,n}} \sum_{b_{(a)}=1}^{\ell_{a,n}} \sum_{s_{(t)}=1}^{\ell_{t,T}} \sum_{v_{(u)}=1}^{\ell_{u,T}} \left| \gamma_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}^{(d_1d_2)} \right| \right) = o(1), \quad (\text{A.4}) \end{aligned}$$

as $T_2/T \rightarrow 0$. Using the similar procedure to (A.4), we can show $D_{3nT} = o(nT\ell_n\ell_T)$, $D_{4nT} = o(nT\ell_n\ell_T)$ and $D_{5nT} = o(nT\ell_n\ell_T)$ given $T_2/T \rightarrow 0$ and $n_2/n \rightarrow 0$.

Thus,

$$\lim_{(n,T) \rightarrow \infty} \frac{1}{nT\ell_n\ell_T} C_{2nT} = \bar{\mathcal{K}}_1 \bar{\mathcal{K}}_2 J(c_1, c_2) J(d_1, d_2).$$

By symmetry,

$$\lim_{(n,T) \rightarrow \infty} \frac{1}{nT\ell_n\ell_T} C_{3nT} = \bar{\mathcal{K}}_1 \bar{\mathcal{K}}_2 J(c_1, d_2) J(c_2, d_1).$$

Therefore,

$$\lim_{(n,T) \rightarrow \infty} \frac{nT}{\ell_n\ell_T} \text{cov} \left(\tilde{J}_{nT}(c_1, d_1), \tilde{J}_{nT}(c_2, d_2) \right) = \bar{\mathcal{K}}_1 \bar{\mathcal{K}}_2 (J(c_1, c_2) J(d_1, d_2) + J(c_1, d_2) J(c_2, d_1)).$$

In terms of matrix form,

$$\lim_{(n,T) \rightarrow \infty} \frac{nT}{\ell_n\ell_T} \text{var} \left(\text{vec} \left(\tilde{J}_{nT} \right) \right) = \bar{\mathcal{K}}_1 \bar{\mathcal{K}}_2 (I_{pp} + \mathbb{K}_{pp}) (J \otimes J).$$

(b) Asymptotic Bias

Let $c_{nT} := d_n^{q_1}/d_T^{q_2}$ and $c_{nT} = c_d + o(1)$ where $c_d > 0$. We have

$$\begin{aligned} d_n^{q_1} \left(E \tilde{J}_{nT} - J_{nT} \right) &= \frac{1}{nT} \sum_{i,j=1}^n \sum_{t,s=1}^T \Gamma_{(it,js)} \left[(d_{ij})^{q_1} \frac{K_1 \left(\frac{d_{ij}}{d_n} \right) - 1}{\left(\frac{d_{ij}}{d_n} \right)^{q_1}} + d_n^{q_1} \left(\frac{d_{ts}}{d_T} \right)^{q_2} \frac{K_2 \left(\frac{d_{ts}}{d_T} \right) - 1}{\left(\frac{d_{ts}}{d_T} \right)^{q_2}} \right. \\ &\quad \left. + (d_{ij})^{q_1} \left(\frac{d_{ts}}{d_T} \right)^{q_2} \frac{\left(K_1 \left(\frac{d_{ij}}{d_n} \right) - 1 \right) \left(K_2 \left(\frac{d_{ts}}{d_T} \right) - 1 \right)}{\left(\frac{d_{ij}}{d_n} \right)^{q_1} \left(\frac{d_{ts}}{d_T} \right)^{q_2}} \right] \\ &= -K_{q_1} b_1^{(q_1)} - c_d K_{q_2} b_2^{(q_2)} + o(1). \end{aligned}$$

Therefore, $\lim_{(n,T) \rightarrow \infty} d_n^{q_1} (\tilde{J}_{nT} - J_{nT}) = -K_{q_1} b_1^{(q_1)} - c_d K_{q_2} b_2^{(q_2)}$.

(c) AMSE

Since

$$\frac{nT}{\ell_n\ell_T} = \frac{d_n^{2q_1}}{d_n^{2q_1} \ell_n\ell_T / nT} = \frac{d_n^{2q_1}}{\tau + o(1)},$$

we have

$$\begin{aligned}
& \lim_{(n,T) \rightarrow \infty} MSE \left(\frac{nT}{\ell_n \ell_T}, \tilde{J}_{nT}, S \right) \\
&= \lim_{(n,T) \rightarrow \infty} \frac{nT}{\ell_n \ell_T} \text{vec} \left(E \tilde{J}_{nT} - J_{nT} \right)' S \text{vec} \left(E \tilde{J}_{nT} - J_{nT} \right) + \lim_{(n,T) \rightarrow \infty} \frac{nT}{\ell_n \ell_T} \bar{\mathcal{K}}_1 \bar{\mathcal{K}}_2 \text{tr} \left(S \text{var}(\text{vec } \tilde{J}_{nT}) \right) \\
&= \frac{1}{\tau} \text{vec} \left(K_{q_1} b_1^{(q_1)} + c_d K_{q_2} b_2^{(q_2)} \right)' S \text{vec} \left(K_{q_1} b_1^{(q_1)} + c_d K_{q_2} b_2^{(q_2)} \right) + \bar{\mathcal{K}}_1 \bar{\mathcal{K}}_2 \text{tr} [S(I_{pp} + \mathbb{K}_{pp})(J \otimes J)],
\end{aligned}$$

where the last equality holds by Theorem 1(a) and (b).

Proof of Corollary 1

For any sequence of (d_n, d_T) , let $\tau_{nT} := \tau_{nT}(d_n, d_T) = d_n^{2q_1} \ell_n \ell_T / (nT)$ and $c_{nT} := c_{nT}(d_n, d_T) := d_n^{q_1} / d_T^{q_2}$. The mapping between (d_n, d_T) and (τ_{nT}, c_{nT}) is one-to-one and invertible. We can express (d_n, d_T) in terms of (τ_{nT}, c_{nT}) as follows:

$$\begin{aligned}
d_n &= (\tau_{nT})^{\frac{q_2}{q_1 \eta_T + 2q_1 q_2 + q_2 \eta_n}} (c_{nT})^{\frac{\eta_T}{q_1 \eta_T + 2q_1 q_2 + q_2 \eta_n}} \left(\frac{nT}{\alpha_n \alpha_T} \right)^{\frac{q_2}{q_1 \eta_T + 2q_1 q_2 + q_2 \eta_n}}, \\
d_T &= (\tau_{nT})^{\frac{q_1}{q_1 \eta_T + 2q_1 q_2 + q_2 \eta_n}} (c_{nT})^{-\frac{\eta_n + 2q_1}{q_1 \eta_T + 2q_1 q_2 + q_2 \eta_n}} \left(\frac{nT}{\alpha_n \alpha_T} \right)^{\frac{q_1}{q_1 \eta_T + 2q_1 q_2 + q_2 \eta_n}}.
\end{aligned}$$

Now

$$\begin{aligned}
& \max_{(b_1^{(q_1)}, b_2^{(q_2)}) \in \mathfrak{B}} AMSE \left((nT)^{\frac{2q_1 q_2}{q_1 \eta_T + 2q_1 q_2 + q_2 \eta_n}}, \hat{J}_{nT}(d_n, d_T), S \right) \\
&= (nT)^{\frac{2q_1 q_2}{q_1 \eta_T + 2q_1 q_2 + q_2 \eta_n}} \left[2 \left(K_{q_1}^2 \frac{B_{11}}{d_n^{2q_1}} + K_{q_2}^2 \frac{B_{22}}{d_T^{2q_2}} \right) + \frac{\ell_n \ell_T}{nT} \bar{\mathcal{K}}_1 \bar{\mathcal{K}}_2 C_V \right] (1 + o(1)) \\
&= (nT)^{\frac{2q_1 q_2}{q_1 \eta_T + 2q_1 q_2 + q_2 \eta_n}} \left[2 \left(K_{q_1}^2 B_{11} + K_{q_2}^2 c_{nT}^2 B_{22} \right) \frac{1}{d_n^{2q_1}} + \frac{\tau_{nT}}{d_n^{2q_1}} \bar{\mathcal{K}}_1 \bar{\mathcal{K}}_2 C_V \right] (1 + o(1)) \\
&= (\alpha_n \alpha_T)^{\frac{2q_1 q_2}{q_1 \eta_T + 2q_1 q_2 + q_2 \eta_n}} (\tau_{nT})^{\frac{-2q_1 q_2}{q_1 \eta_T + 2q_1 q_2 + q_2 \eta_n}} (c_{nT})^{\frac{-2q_1 \eta_T}{q_1 \eta_T + 2q_1 q_2 + q_2 \eta_n}} \\
&\quad \times \left[2 \left(K_{q_1}^2 B_{11} + K_{q_2}^2 c_{nT}^2 B_{22} \right) + \tau_{nT} \bar{\mathcal{K}}_1 \bar{\mathcal{K}}_2 C_V \right] (1 + o(1)).
\end{aligned}$$

Some elementary calculations show that the above dominating term is uniquely minimized over $(\tau_{nT}, c_{nT}) \in \mathbb{R}_+^2$ at (τ_{nT}^*, c_{nT}^*) where

$$\tau_{nT}^* = (d_n^*)^{2q_1} \ell_n \ell_T / (nT) \text{ and } c_{nT}^* = (d_n^*)^{q_1} / (d_T^*)^{q_2},$$

and $\tau_{nT}^* \rightarrow \tau^* \in (0, \infty)$ and $c_{nT}^* \rightarrow c_d^* \in (0, \infty)$. As a result

$$\begin{aligned}
& \lim_{(n,T) \rightarrow \infty} \max_{(b_1^{(q_1)}, b_2^{(q_2)}) \in \mathfrak{B}} AMSE \left((nT)^{\frac{2q_1 q_2}{q_1 \eta_T + 2q_1 q_2 + q_2 \eta_n}}, \hat{J}_{nT}(d_n, d_T), S \right) \\
&\geq \lim_{(n,T) \rightarrow \infty} \max_{(b_1^{(q_1)}, b_2^{(q_2)}) \in \mathfrak{B}} AMSE \left((nT)^{\frac{2q_1 q_2}{q_1 \eta_T + 2q_1 q_2 + q_2 \eta_n}}, \hat{J}_{nT}(d_n^*, d_T^*), S \right). \tag{A.5}
\end{aligned}$$

The inequality holds with equality if and only if $\tau_{nT} = \tau_{nT}^* (1 + o(1))$ and $c_{nT} = c_{nT}^* (1 + o(1))$. In other words, the inequality is strict unless $d_n = d_n^* (1 + o(1))$ and $d_T = d_T^* (1 + o(1))$.

Proof of Proposition 1

(a) $\hat{J}_{nT} - \hat{J}_{nT}^{GA} = o_p(1)$ if $d_n \rightarrow 0$ as $n \rightarrow \infty$.

Under the assumptions of Proposition 1, it is enough to show that

$$\tilde{J}_{nT}(c, d) - \tilde{J}_{nT}^{GA}(c, d) = o_p(1). \quad (\text{A.6})$$

Recall $V_{i,t}^{(c)} = \sum_{l=1}^{nTp} r_{(i,t),l}^{(c)} \varepsilon_l$. By Chebyshev's inequality, for any $\Delta > 0$,

$$\begin{aligned} & P\left(\left|\tilde{J}_{nT}(c, d) - \tilde{J}_{nT}^{GA}(c, d)\right| > \Delta\right) \\ & \leq \frac{1}{\Delta^2} \frac{1}{n^2 T^2} \sum_{i \neq j}^T \sum_{t,s=1}^T \sum_{a \neq b}^T \sum_{u,v=1}^T K_1\left(\frac{d_{ij}}{d_n}\right) K_1\left(\frac{d_{ab}}{d_n}\right) K_2\left(\frac{d_{ts}}{d_T}\right) K_2\left(\frac{d_{uv}}{d_T}\right) E\left[V_{(i,t)}^{(c)} V_{(j,s)}^{(d)} V_{(a,u)}^{(c)} V_{(b,v)}^{(d)}\right] \\ & := \frac{1}{\Delta^2} \frac{1}{n^2 T^2} \left(\tilde{C}_{1nT} + \tilde{C}_{2nT} + \tilde{C}_{3nT} + \tilde{C}_{4nT}\right), \end{aligned}$$

where

$$\begin{aligned} \tilde{C}_{1nT} &= \sum_{t,s,u,v=1}^T \sum_{l=1}^{nTp} \sum_{i \neq j} \sum_{a \neq b} K_1\left(\frac{d_{ij}}{d_n}\right) K_1\left(\frac{d_{ab}}{d_n}\right) K_2\left(\frac{d_{ts}}{d_T}\right) K_2\left(\frac{d_{uv}}{d_T}\right) r_{(i,t),l}^{(c)} r_{(j,s),l}^{(d)} r_{(a,u),l}^{(c)} r_{(b,v),l}^{(d)} (E\varepsilon_l^4 - 3) \\ \tilde{C}_{2nT} &= \sum_{t,s,u,v=1}^T \sum_{l,k=1}^{nTp} \sum_{i \neq j} \sum_{a \neq b} K_1\left(\frac{d_{ij}}{d_n}\right) K_1\left(\frac{d_{ab}}{d_n}\right) K_2\left(\frac{d_{ts}}{d_T}\right) K_2\left(\frac{d_{uv}}{d_T}\right) r_{(i,t),l}^{(c)} r_{(j,s),l}^{(d)} r_{(a,u),k}^{(c)} r_{(b,v),k}^{(d)} \\ \tilde{C}_{3nT} &= \sum_{t,s,u,v=1}^T \sum_{l,k=1}^{nTp} \sum_{i \neq j} \sum_{a \neq b} K_1\left(\frac{d_{ij}}{d_n}\right) K_1\left(\frac{d_{ab}}{d_n}\right) K_2\left(\frac{d_{ts}}{d_T}\right) K_2\left(\frac{d_{uv}}{d_T}\right) r_{(i,t),l}^{(c)} r_{(j,s),k}^{(d)} r_{(a,u),l}^{(c)} r_{(b,v),k}^{(d)} \\ \tilde{C}_{4nT} &= \sum_{t,s,u,v=1}^T \sum_{l,k=1}^{nTp} \sum_{i \neq j} \sum_{a \neq b} K_1\left(\frac{d_{ij}}{d_n}\right) K_1\left(\frac{d_{ab}}{d_n}\right) K_2\left(\frac{d_{ts}}{d_T}\right) K_2\left(\frac{d_{uv}}{d_T}\right) r_{(i,t),l}^{(c)} r_{(j,s),k}^{(d)} r_{(a,u),k}^{(c)} r_{(b,v),l}^{(d)}. \end{aligned}$$

Using the same argument as in (A.1), we can show $(nT)^{-2} \tilde{C}_{1nT} = o(n^2 T^2)$. For \tilde{C}_{2nT} ,

$$\frac{1}{\Delta^2} \frac{1}{n^2 T^2} \tilde{C}_{2nT} \leq \frac{1}{\Delta^2} \left(\frac{1}{nT} \sum_{t,s=1}^T \sum_{i \neq j} K_1\left(\frac{d_{ij}}{d_n}\right) \left| \gamma_{(it,js)}^{(cd)} \right| \right)^2 \rightarrow 0$$

as $d_n \rightarrow 0$ because $K_1(d_{ij}/d_n) = 0$ for all $i \neq j$ provided $d_n < \min_{i,j} d_{ij}$. With the similar procedures, we can show that $\tilde{C}_{3nT} \rightarrow 0$ and $\tilde{C}_{4nT} \rightarrow 0$. Therefore, (A.6) holds.

(b) $\hat{J}_{nT} - \hat{J}_{nT}^{DK} = o_p(1)$ if $\ell_n/n \rightarrow 1$ as $n \rightarrow \infty$.

It suffices to show that

$$\tilde{J}_{nT}(c, d) - \tilde{J}_{nT}^{DK}(c, d) = o_p(1). \quad (\text{A.7})$$

Note that

$$\tilde{J}_{nT}(c, d) - \tilde{J}_{nT}^{DK}(c, d) = \frac{1}{nT} \sum_{i,j=1}^n \sum_{t,s=1}^T \left[K_1\left(\frac{d_{ij}}{d_n}\right) - 1 \right] K_2\left(\frac{d_{ts}}{d_T}\right) V_{(i,t)}^{(c)} V_{(j,s)}^{(d)}.$$

By Chebyshev's inequality, we have

$$\begin{aligned} P\left(\left|\tilde{J}_{nT}(c, d) - \tilde{J}_{nT}^{DK}(c, d)\right| > \Delta\right) &\leq \frac{1}{\Delta^2} E\left(\tilde{J}_{nT}(c, d) - \tilde{J}_{nT}^{DK}(c, d)\right)^2 \\ &:= \check{C}_{1nT} + \check{C}_{2nT} + \check{C}_{3nT} + \check{C}_{4nT}, \end{aligned}$$

for any Δ , where

$$\begin{aligned} \check{C}_{1nT} &= \frac{1}{\Delta^2} \frac{1}{n^2 T^2} \sum_{i,j,a,b=1}^n \sum_{t,s,u,v=1}^T \sum_{l=1}^{nTp} \left(K_1\left(\frac{d_{ij}}{d_n}\right) - 1\right) \left(K_1\left(\frac{d_{ab}}{d_n}\right) - 1\right) K_2\left(\frac{d_{ts}}{d_T}\right) K_2\left(\frac{d_{uv}}{d_T}\right) \\ &\quad \times r_{(i,t),l}^{(c)} r_{(j,s),l}^{(d)} r_{(a,u),l}^{(c)} r_{(b,v),l}^{(d)} (E\varepsilon_l^4 - 3) \\ \check{C}_{2nT} &= \frac{1}{\Delta^2} \frac{1}{n^2 T^2} \sum_{i,j,a,b=1}^n \sum_{t,s,u,v=1}^T \sum_{l,k=1}^{nTp} \left(K_1\left(\frac{d_{ij}}{d_n}\right) - 1\right) \left(K_1\left(\frac{d_{ab}}{d_n}\right) - 1\right) K_2\left(\frac{d_{ts}}{d_T}\right) K_2\left(\frac{d_{uv}}{d_T}\right) \\ &\quad \times r_{(i,t),l}^{(c)} r_{(j,s),l}^{(d)} r_{(a,u),k}^{(c)} r_{(b,v),k}^{(d)} \\ \check{C}_{3nT} &= \frac{1}{\Delta^2} \frac{1}{n^2 T^2} \sum_{i,j,a,b=1}^n \sum_{t,s,u,v=1}^T \sum_{l,k=1}^{nTp} \left(K_1\left(\frac{d_{ij}}{d_n}\right) - 1\right) \left(K_1\left(\frac{d_{ab}}{d_n}\right) - 1\right) K_2\left(\frac{d_{ts}}{d_T}\right) K_2\left(\frac{d_{uv}}{d_T}\right) \\ &\quad \times r_{(i,t),l}^{(c)} r_{(j,s),k}^{(d)} r_{(a,u),l}^{(c)} r_{(b,v),k}^{(d)} \\ \check{C}_{4nT} &= \frac{1}{\Delta^2} \frac{1}{n^2 T^2} \sum_{i,j,a,b=1}^n \sum_{t,s,u,v=1}^T \sum_{l,k=1}^{nTp} \left(K_1\left(\frac{d_{ij}}{d_n}\right) - 1\right) \left(K_1\left(\frac{d_{ab}}{d_n}\right) - 1\right) K_2\left(\frac{d_{ts}}{d_T}\right) K_2\left(\frac{d_{uv}}{d_T}\right) \\ &\quad \times r_{(i,t),l}^{(c)} r_{(j,s),k}^{(d)} r_{(a,u),k}^{(c)} r_{(b,v),l}^{(d)}. \end{aligned}$$

We can show that $\check{C}_{1nT} = o(1)$ using the same argument as in (A.1). For \check{C}_{2nT} , as $K_1(\cdot)$ is the rectangular kernel,

$$\begin{aligned} \check{C}_{2nT} &= \frac{1}{\Delta^2} \frac{1}{n^2 T^2} \sum_{i,j,a,b=1}^n \sum_{t,s,u,v=1}^T \left[K_1\left(\frac{d_{ij}}{d_n}\right) - 1 \right] \left[K_1\left(\frac{d_{ab}}{d_n}\right) - 1 \right] K_2\left(\frac{d_{ts}}{d_T}\right) K_2\left(\frac{d_{uv}}{d_T}\right) \gamma_{(it,jt)}^{(cd)} \gamma_{(au,bv)}^{(cd)} \\ &\leq \frac{1}{\Delta^2} \left(\frac{1}{nT} \sum_{i,j=1}^n \sum_{t,s=1}^T \left[K_1\left(\frac{d_{ij}}{d_n}\right) - 1 \right] d_{ij}^{-qs} \left| \gamma_{(it,jt)}^{(cd)} \right| d_{ij}^{qs} \right)^2 \\ &= \frac{1}{\Delta^2} \left(\frac{1}{nT} \sum_{i,j=1}^n \sum_{t,s=1}^T 1 \left\{ \frac{d_{ij}}{d_n} > 1 \right\} d_{ij}^{-qs} \left| \gamma_{(it,jt)}^{(cd)} \right| d_{ij}^{qs} \right)^2 \\ &\leq \frac{1}{\Delta^2} \left(\frac{1}{nT} \sum_{i,j=1}^n \sum_{t,s=1}^T 1 \left\{ \frac{d_{ij}}{d_n} > 1 \right\} (d_n)^{-qs} \left| \gamma_{(it,jt)}^{(cd)} \right| d_{ij}^{qs} \right)^2 \\ &\leq \frac{1}{\Delta^2} (d_n)^{-2qs} \left(\frac{1}{nT} \sum_{i,j=1}^n \sum_{t,s=1}^T \left| \gamma_{(it,jt)}^{(cd)} \right| d_{ij}^{qs} \right)^2 \\ &\rightarrow 0 \end{aligned}$$

as $d_n \rightarrow \infty$. For \check{C}_{3nT} , we have:

$$\begin{aligned}
\check{C}_{3nT} &= \frac{1}{\Delta^2} \frac{1}{n^2 T^2} \sum_{i,j,a,b=1}^n \sum_{t,s,u,v=1}^T \left(K_1 \left(\frac{d_{ij}}{d_n} \right) - 1 \right) \left(K_1 \left(\frac{d_{ab}}{d_n} \right) - 1 \right) \\
&\quad \times K_2 \left(\frac{d_{ts}}{d_T} \right) K_2 \left(\frac{d_{uv}}{d_T} \right) \gamma_{(it,au)}^{(cc)} \gamma_{(js,bv)}^{(dd)} \\
&\leq \frac{1}{\Delta^2} \frac{1}{n^2 T^2} \sum_{i,j,a,b=1}^n \sum_{t,s,u,v=1}^T \left(K_1 \left(\frac{d_{ij}}{d_n} \right) - 1 \right) \left(K_1 \left(\frac{d_{ab}}{d_n} \right) - 1 \right) \left| \gamma_{(it,au)}^{(cc)} \gamma_{(js,bv)}^{(dd)} \right| \\
&= \frac{1}{\Delta^2} \frac{1}{n^2 T^2} \sum_{i,j,a,b=1}^n \sum_{t,s,u,v=1}^T 1 \left\{ \frac{d_{ij}}{d_n} > 1 \right\} 1 \left\{ \frac{d_{ab}}{d_n} > 1 \right\} \left| \gamma_{(it,au)}^{(cc)} \gamma_{(js,bv)}^{(dd)} \right| \\
&= \frac{1}{\Delta^2} \frac{1}{n^2 T^2} \sum_{i,j,a,b=1}^n \sum_{t,s,u,v=1}^T 1 \left\{ \frac{d_{ij}}{d_n} > 1 \right\} 1 \left\{ \frac{d_{ab}}{d_n} > 1 \right\} 1 \left\{ \frac{d_{ia}}{d_n} \leq 1 \right\} 1 \left\{ \frac{d_{jb}}{d_n} \leq 1 \right\} \\
&\quad \times \left| \gamma_{(it,au)}^{(cc)} \gamma_{(js,bv)}^{(dd)} \right| + o(1) \\
&= \frac{1}{\Delta^2} \frac{1}{nT} \sum_{i=1}^n \sum_{a: d_{ia}/d_n \leq 1} \sum_{t=1}^T \sum_{u=1}^T \left| \gamma_{(it,au)}^{(cc)} \right| \left(\frac{1}{nT} \sum_{j: d_{ij}/d_n > 1} \sum_{\{b: d_{jb}/d_n \leq 1, d_{ab}/d_n > 1\}} \sum_{s=1}^T \sum_{v=1}^T \left| \gamma_{(js,bv)}^{(dd)} \right| \right) + o(1).
\end{aligned}$$

As $\ell_{i,n} \leq c\ell_n$ with some constant c , if $\ell_n/n \rightarrow 1$, then

$$\begin{aligned}
&\frac{1}{nT} \sum_{\{j: d_{ij}/d_n > 1\}} \sum_{\{b: d_{jb}/d_n \leq 1, d_{ab}/d_n > 1\}} \sum_{s,v=1}^T \left| \gamma_{(js,bv)}^{(dd)} \right| \\
&= \frac{n - \ell_n}{n} \frac{1}{(n - \ell_n)T} \sum_{\{j: d_{ij}/d_n > 1\}} \sum_{\{b: d_{jb}/d_n \leq 1, d_{ab}/d_n > 1\}} \sum_{s,v=1}^T \left| \gamma_{(js,bv)}^{(dd)} \right| \rightarrow 0,
\end{aligned}$$

which implies $\check{C}_{3nT} \rightarrow 0$ as $(n, T) \rightarrow \infty$. With the same procedure, we can show that $\check{C}_{4nT} = o(1)$. Therefore, (A.7) holds.

(c) $\hat{J}_{nT} - \hat{J}_{nT}^{DK*} = o_p(1)$ if $\ell_T/T \rightarrow 1$ as $T \rightarrow \infty$.

The proof is analogous to the proof of (b).

The proof of Proposition 2 uses the lemma below whose proof is given in the supplementary appendix.

Lemma 1 *Let*

$$\mathfrak{X}_{\iota,nT} := \frac{1}{\sqrt{nT}} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^T \Phi_{b,\iota} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) \hat{V}_{(i_1,i_2,t)}^*.$$

Then, under Assumptions F3 – F4

$$\mathfrak{X}_{\iota,nT} \stackrel{a}{\sim} \mathfrak{X}_{\iota,nT}^a := \Lambda \frac{1}{\sqrt{nT}} \sum_{i_1,i_2,t} \Phi_{b,\iota} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) \left(e_{(i_1,i_2,t)}^* - \frac{1}{nT} \sum_{i_1,i_2,t} e_{(i_1,i_2,t)}^* \right)$$

jointly over $\iota = 1, 2, \dots, \mathcal{L}$ for every fixed \mathcal{L} .

Lemma 2 Suppose $\zeta_{nT} = \xi_{nT,\mathcal{L}} + \eta_{nT,\mathcal{L}}$. Assume

- (i) $P(\xi_{nT,\mathcal{L}} < \xi) - P(\xi_{nT,\mathcal{L}}^* < \xi) = o(1)$ for each fixed \mathcal{L} and each $\xi \in \mathbb{R}$ as $(n, T) \rightarrow \infty$,
- (ii) $P(\xi_{nT,\mathcal{L}}^* < \xi) - P(\xi_{nT} < \xi) = o(1)$ uniformly in (n, T) for each $\xi \in \mathbb{R}$ as $\mathcal{L} \rightarrow \infty$,
- (iii) The CDF of ξ_{nT} is equicontinuous when n and T are sufficiently large,
- (iv) $\eta_{nT,\mathcal{L}} \xrightarrow{p} 0$ uniformly in (n, T) as $\mathcal{L} \rightarrow \infty$. Then

$$P(\zeta_{nT} < \xi) = P(\xi_{nT} < \xi) + o(1) \text{ for each } \xi \in \mathbb{R} \text{ as } (n, T) \rightarrow \infty.$$

Proof of Lemma 2. Let $\varepsilon > 0$. Under condition (iii), we can find $\delta > 0$ such that for some integer $C > 0$,

$$P(\xi - \delta \leq \xi_{nT} < \xi + \delta) \leq \varepsilon$$

for all (n, T) with $\min(n, T) \geq C$. Under condition (iv), we can find an N such that

$$P(|\eta_{nT,\mathcal{L}}| > \delta) \leq \varepsilon$$

for all $\mathcal{L} \geq N$ and all (n, T) . From condition (ii), we can find $N' \geq N$ such that

$$|P(\xi_{nT,\mathcal{L}}^* < \xi) - P(\xi_{nT} < \xi)| \leq \varepsilon$$

for all $\mathcal{L} \geq N'$ and all n and T . It follows from condition (i) that for any fixed $\mathcal{L}_0 \geq N'$, there exists a $\tilde{C}(\mathcal{L}_0) \geq C$ such that

$$|P(\xi_{nT,\mathcal{L}_0} < \xi) - P(\xi_{nT,\mathcal{L}_0}^* < \xi)| \leq \varepsilon$$

for (n, T) with $\min(n, T) \geq \tilde{C}(\mathcal{L}_0)$.

When $\min(n, T) \geq \tilde{C}(\mathcal{L}_0)$, we have

$$\begin{aligned} P(\zeta_{nT} \leq \xi) &= P(\xi_{nT,\mathcal{L}_0} + \eta_{nT,\mathcal{L}_0} \leq \xi) \\ &\leq P(\xi_{nT,\mathcal{L}_0} \leq \xi + \delta) + P(|\eta_{nT,\mathcal{L}_0}| > \delta) \\ &\leq P(\xi_{nT,\mathcal{L}_0}^* \leq \xi + \delta) + 2\varepsilon \leq P(\xi_{nT} < \xi + \delta) + 3\varepsilon \\ &\leq P(\xi_{nT} < \xi) + 4\varepsilon. \end{aligned}$$

Similarly,

$$\begin{aligned} P(\zeta_{nT} \leq \xi) &= P(\xi_{nT,\mathcal{L}_0} + \eta_{nT,\mathcal{L}_0} \leq \xi) \\ &\geq P(\xi_{nT,\mathcal{L}_0} \leq \xi - \delta) - P(|\eta_{nT,\mathcal{L}_0}| \geq \delta) \\ &\geq P(\xi_{nT,\mathcal{L}_0}^* \leq \xi - \delta) - 2\varepsilon \geq P(\xi_{nT} \leq \xi - \delta) - 3\varepsilon \\ &\geq P(\xi_{nT} \leq \xi) - 4\varepsilon. \end{aligned}$$

Since the above two inequalities hold for all $\varepsilon > 0$, we must have $P(\zeta_{nT} < \xi) = P(\xi_{nT} < \xi) + o(1)$.

Proof of Proposition 2

We first consider the case that $K_b(\cdot, \cdot)$ is continuous on $([0, 1]^3 \times [0, 1]^3)$ and continuously differentiable almost everywhere on $([0, 1]^3 \times [0, 1]^3)$. Let

$$\check{J}_{nT} = \frac{1}{nT} \sum_{i_1, j_1=1}^{L_n} \sum_{i_2, j_2=1}^{M_n} \sum_{t,s=1}^T \sum_{\iota=1}^{\infty} \lambda_{\iota} \Phi_{b,\iota} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) \Phi_{b,\iota} \left(\frac{j_1}{L_n}, \frac{j_2}{M_n}, \frac{s}{T} \right) \hat{V}_{(i_1, i_2, t)}^* \hat{V}_{(j_1, j_2, s)}^{*'}.$$

Since

$$\mathbb{K}_b((x_1, x_2, x_3), (y_1, y_2, y_3)) = \lim_{\mathcal{L} \rightarrow \infty} \sum_{\iota=1}^{\mathcal{L}} \lambda_{\iota} \Phi_{b,\iota}((x_1, x_2, x_3)) \Phi_{b,\iota}(y_1, y_2, y_3), \quad (\text{A.8})$$

where the right hand side converges absolutely and uniformly over $([0, 1]^3 \times [0, 1]^3)$, \check{J}_{nT} and \hat{J}_{nT} have the same limiting distribution.

We use Lemma 1 and Lemma 2 to complete the proof. We write

$$\check{J}_{nT} = \xi_{nT,\mathcal{L}} + \eta_{nT,\mathcal{L}} \text{ for } \xi_{nT,\mathcal{L}} = \sum_{\iota=1}^{\mathcal{L}} \lambda_{\iota} \mathfrak{X}_{\iota,nT} \mathfrak{X}'_{\iota,nT} \text{ and } \eta_{nT,\mathcal{L}} = \sum_{\iota=\mathcal{L}+1}^{\infty} \lambda_{\iota} \mathfrak{X}_{\iota,nT} \mathfrak{X}'_{\iota,nT}.$$

By Lemma 1, for each fixed \mathcal{L} ,

$$\xi_{nT,\mathcal{L}} \stackrel{a}{\sim} \xi_{nT,\mathcal{L}}^* := \sum_{\iota=1}^{\mathcal{L}} \lambda_{\iota} \mathfrak{X}_{\iota,nT}^a \mathfrak{X}_{\iota,nT}^{a'}$$

so condition (i) in Lemma 2 is satisfied. Let $\xi_{nT} := \sum_{\iota=1}^{\infty} \lambda_{\iota} \mathfrak{X}_{\iota,nT}^a \mathfrak{X}_{\iota,nT}^{a'}$, then it is easy to see that condition (ii) in Lemma 2 holds. The uniformity in condition (ii) holds because

$$\xi_{nT} - \xi_{nT,\mathcal{L}}^* = \sum_{\iota=\mathcal{L}+1}^{\infty} \lambda_{\iota} \mathfrak{X}_{\iota,nT}^a \mathfrak{X}_{\iota,nT}^{a'}$$

and for any r_1 and $r_2 \in \mathbb{R}^p$

$$E |r_1' (\xi_{nT} - \xi_{nT,\mathcal{L}}^*) r_2| \leq C \sum_{\iota=\mathcal{L}+1}^{\infty} |\lambda_{\iota}| \rightarrow 0 \quad (\text{A.9})$$

uniformly in (n, T) for some constant $C > 0$. To verify condition (iii) in Lemma 2, we note that

$$\mathfrak{X}_{\iota,nT}^a \rightarrow^d \mathfrak{X}_{\iota}^a$$

jointly for $\iota = 1, 2, \dots, \mathcal{L}_{\mathfrak{X}}$ for any fixed constant $\mathcal{L}_{\mathfrak{X}}$, where $\{\mathfrak{X}_{\iota}^a, \iota = 1, 2, \dots, \mathcal{L}_{\mathfrak{X}}\}$ are jointly normal with

$$\begin{aligned} \text{var}(\mathfrak{X}_{\iota}^a) &= J \lim_{(n,T) \rightarrow \infty} \frac{1}{nT} \sum_{i_1, i_2, t} \left[\Phi_{b,\iota} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) 1_{i_1, i_2} - \bar{\Phi}_{b,\iota} \right]^2, \\ \text{cov}(\mathfrak{X}_{\iota_1}^a, \mathfrak{X}_{\iota_2}^a) &= J \lim_{(n,T) \rightarrow \infty} \frac{1}{nT} \sum_{i_1, i_2, t} \left[\Phi_{b,\iota_1} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) 1_{i_1, i_2} - \bar{\Phi}_{b,\iota_1} \right] \left[\Phi_{b,\iota_2} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) 1_{i_1, i_2} - \bar{\Phi}_{b,\iota_2} \right]. \end{aligned}$$

In the above, $\bar{\Phi}_{b,\iota} = (nT)^{-1} \sum_{i_1, i_2, t} 1_{i_1, i_2} \Phi_{b,\iota}(i_1/L_n, i_2/M_n, t/T)$. Now

$$\xi_{nT} := \sum_{\iota=1}^{\mathcal{L}_{\mathfrak{X}}} \lambda_{\iota} \mathfrak{X}_{\iota,nT}^a \mathfrak{X}_{\iota,nT}^{a'} + \sum_{\iota=\mathcal{L}_{\mathfrak{X}}+1}^{\infty} \lambda_{\iota} \mathfrak{X}_{\iota,nT}^a \mathfrak{X}_{\iota,nT}^{a'} := \xi_{nT,\mathcal{L}_{\mathfrak{X}}}^{(1)} + \xi_{nT,\mathcal{L}_{\mathfrak{X}}}^{(2)}.$$

Using the joint distributional convergence of $\{\mathfrak{X}_{\iota,nT}^a\}$ to $\{\mathfrak{X}_{\iota}^a\}$, we have $\xi_{nT,\mathcal{L}_{\mathfrak{X}}}^{(1)} \rightarrow^d \sum_{\iota=1}^{\mathcal{L}_{\mathfrak{X}}} \lambda_{\iota} \mathfrak{X}_{\iota}^a \mathfrak{X}_{\iota}^{a'}$ which in turn converges to $\xi_{\infty} := \sum_{\iota=1}^{\infty} \lambda_{\iota} \mathfrak{X}_{\iota}^a \mathfrak{X}_{\iota}^{a'}$ as $\mathcal{L}_{\mathfrak{X}} \rightarrow \infty$. On the other hand,

$$\xi_{nT,\mathcal{L}_{\mathfrak{X}}}^{(2)} \rightarrow^p 0 \text{ uniformly in } (n, T) \text{ as } \mathcal{L}_{\mathfrak{X}} \rightarrow \infty \quad (\text{A.10})$$

Using the similar argument for proving Lemma 2, we have $\xi_{nT} \rightarrow^d \xi_\infty$, which has a continuous distribution. So for any ξ and δ , we have

$$P(\xi - \delta \leq \xi_{nT} < \xi + \delta) = P(\xi - \delta \leq \xi_\infty < \xi + \delta) + o(1)$$

as $(n, T) \rightarrow \infty$. Given the continuity of the CDF of ξ_∞ , for any $\varepsilon > 0$, we can find a $\delta > 0$ such that $P(\xi - \delta \leq \xi_{nT} < \xi + \delta) \leq \varepsilon$ for all (n, T) when n and T are sufficiently large. We have thus verified Condition (iii) in Lemma 2. Finally, for any r_1 and $r_2 \in \mathbb{R}^p$, we have

$$E|r'_1 \eta_{nT, \mathcal{L}} r_2| \leq \sum_{\iota=\mathcal{L}+1}^{\infty} |\lambda_\iota| E(r'_1 \mathfrak{X}_{\iota, nT} \mathfrak{X}'_{\iota, nT} r_2) \leq C \sum_{\iota=\mathcal{L}+1}^{\infty} |\lambda_\iota| \rightarrow 0 \quad (\text{A.11})$$

uniformly in (n, T) as $\mathcal{L} \rightarrow \infty$. Hence $\eta_{nT, \mathcal{L}} \rightarrow^p 0$ uniformly in (n, T) as $\mathcal{L} \rightarrow \infty$.

It follows from the above steps that

$$\begin{aligned} \check{J}_{nT} &\stackrel{a}{\sim} \sum_{\iota=1}^{\infty} \lambda_\iota \mathfrak{X}_{\iota, nT}^a \mathfrak{X}_{\iota, nT}^{a'} \\ &\stackrel{d}{=} \Lambda \frac{1}{nT} \sum_{i_1, j_1, i_2, j_2, t, s} \mathbb{K}_b \left(\left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right), \left(\frac{j_1}{L_n}, \frac{j_2}{M_n}, \frac{s}{T} \right) \right) (e_{(i_1, i_2, t)}^* - \bar{e}^*) (e_{(j_1, j_2, s)}^* - \bar{e}^*)' \Lambda', \end{aligned} \quad (\text{A.12})$$

and $\hat{J}_{nT} \stackrel{a}{\sim} \Lambda J_{nT}^a \Lambda'$ as desired.

Next, we consider the case with the rectangular kernel. We follow the same arguments as above but with some modifications. We define

$$\tilde{\beta} = \beta_0 + Q^{-1} \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \tilde{Z}_{it} u_{it},$$

which is asymptotically equivalent to $\hat{\beta}$ in the sense that $\sqrt{nT}(\hat{\beta} - \beta_0) = \sqrt{nT}(\tilde{\beta} - \beta_0) + o_p(1)$.

By construction $\tilde{\beta}$ has a finite fourth moment while $\hat{\beta}$ may not have any moment. On the basis of $\tilde{\beta}$, we introduce another pseudo estimator:

$$\check{J}_{nT} = \frac{1}{nT} \sum_{i_1, j_1, i_2, j_2, t, s} \mathbb{K}_b \left(\left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right), \left(\frac{j_1}{L_n}, \frac{j_2}{M_n}, \frac{s}{T} \right) \right) \tilde{V}_{(i_1, i_2, t)}^* \tilde{V}_{(j_1, j_2, s)}^{*'}$$

and redefine \check{J}_{nT} to be

$$\check{J}_{nT} = \frac{1}{nT} \sum_{i_1, j_1, i_2, j_2, t, s} \sum_{\iota=1}^{\infty} \lambda_\iota \Phi_{b, \iota} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) \Phi_{b, \iota} \left(\frac{j_1}{L_n}, \frac{j_2}{M_n}, \frac{s}{T} \right) \tilde{V}_{(i_1, i_2, t)}^* \tilde{V}_{(j_1, j_2, s)}^{*'}.$$

where $\tilde{V}_{(i_1, i_2, t)}^* = 1_{i_1, i_2} \tilde{V}_{(i_1, i_2, t)}$ and $\tilde{V}_{(i_1, i_2, t)} = \tilde{Z}_{it}(\tilde{Y}_{it} - \tilde{X}_{it}' \tilde{\beta})$. It is not hard to show that $\hat{J}_{nT} = \check{J}_{nT} + o_p(1)$. In addition,

$$\begin{aligned} &\frac{1}{\sqrt{nT}} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^T \Phi_{b, \iota} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) \tilde{V}_{(i_1, i_2, t)}^* \\ &= \frac{1}{\sqrt{nT}} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^T \Phi_{b, \iota} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) \tilde{V}_{(i_1, i_2, t)}^* + o_p(1) \end{aligned}$$

for $\iota = 1, 2, \dots, \mathcal{L}$. So Lemma 1 holds with $\hat{V}_{(i_1, i_2, t)}^*$ replaced by $\tilde{V}_{(i_1, i_2, t)}^*$.

It remains to show that $\check{J}_{nT} = \check{J}_{nT} + o_p(1)$ and \check{J}_{nT} is asymptotically equivalent in distribution to $\Lambda J_{nT}^a \Lambda'$. For the first result, we can use the L_1 and L_2 convergence of the Fourier series expansion and show via some tedious calculations that $E \left[\text{vec}(\check{J}_{nT} - \check{J}_{nT})' \text{Svec}(\check{J}_{nT} - \check{J}_{nT}) \right] = o(1)$ as $(n, T) \rightarrow \infty$. For the second result, we use the similar argument for proving (A.12). We only have to modify (A.9), (A.10) and (A.11). Instead of using the absolute moment as in (A.9) and (A.11) to obtain the convergence in probability, we compute the mean and variance and use Chebyshev's inequality to obtain the same result. The modifications are needed because when the kernel functions are continuous and piecewise smooth, we have $\sum_{\iota=1}^{\infty} |\lambda_{\iota}| < \infty$. In contrast, for the rectangular kernel, we have only $\sum_{\iota=1}^{\infty} \lambda_{\iota} < \infty$ and $\sum_{\iota=1}^{\infty} \lambda_{\iota}^2 < \infty$.

Lemma 3 *As $(b_1, b_2, b_3) \rightarrow 0$, we have*

$$(a) \mu_1 = 1 - b_1 b_2 b_3 c_1 + o(b_1 b_2 b_3); (b) \mu_2 = b_1 b_2 b_3 c_2 + o(b_1 b_2 b_3).$$

Proofs are given in the supplementary appendix.

Lemma 4 *As $(b_1, b_2, b_3) \rightarrow 0$, we have*

$$\begin{aligned} (a) E(v_{11} - v_{12} v_{22}^{-1} v_{21}) &= 1 - b_1 b_2 b_3 c_1 - (g-1) b_1 b_2 b_3 c_2 + o(b_1 b_2 b_3), \\ (b) E(v_{11} - v_{12} v_{22}^{-1} v_{21})^2 &= 1 - 2b_1 b_2 b_3 (c_1 + (g-2) c_2) + o(b_1 b_2 b_3), \\ (c) E[(v_{11} - v_{12} v_{22}^{-1} v_{21}) - 1]^2 &= 2b_1 b_2 b_3 c_2 + o(b_1 b_2 b_3). \end{aligned}$$

Proof of Lemma 4

This is a direct application of Lemma 3 in Sun (2010).

Proof of Theorem 6

Taking a Taylor expansion, we have

$$\begin{aligned} P\{gF_{\infty}(g, b) \leq z\} &= EG_g(z(v_{11} - v_{12} v_{22}^{-1} v_{21})) \\ &= G_g(z) + G'_g(z) z E[(v_{11} - v_{12} v_{22}^{-1} v_{21}) - 1] + \frac{1}{2} G''_g(z) z^2 E[(v_{11} - v_{12} v_{22}^{-1} v_{21}) - 1]^2 \\ &\quad + \frac{1}{2} E[G''_g(\tilde{z}) - G''_g(z)] z^2 [(v_{11} - v_{12} v_{22}^{-1} v_{21}) - 1]^2 \end{aligned}$$

where \tilde{z} is between z and $z(v_{11} - v_{12} v_{22}^{-1} v_{21})$. Using Lemma 4, we have

$$\begin{aligned} P\{gF_{\infty}(g, b) \leq z\} &= G_g(z) - G'_g(z) z [b_1 b_2 b_3 c_1 + (g-1) b_1 b_2 b_3 c_2] + G''_g(z) z^2 b_1 b_2 b_3 c_2 + o(b_1 b_2 b_3) \\ &= G_g(z) + [G''_g(z) z^2 c_2 - G'_g(z) z (c_1 + (g-1) c_2)] b_1 b_2 b_3 + o(b_1 b_2 b_3) \\ &= G_g(z) + A(z) b_1 b_2 b_3 + o(b_1 b_2 b_3). \end{aligned}$$

Proof of Theorem 7

Using Lemma 3, we have

$$D = \frac{1}{b_1 b_2 b_3 c_2} (1 + o(1)), \quad (\text{A.13})$$

$$\frac{\mu_1(D - g + 1)}{D} = \frac{1}{1 + b_1 b_2 b_3 [c_1 + (g - 1) c_2]} + o(b_1 b_2 b_3). \quad (\text{A.14})$$

It now follows from (A.14) and Theorem 6 that

$$\begin{aligned} & P \left\{ \frac{\mu_1(D - g + 1)}{D} F_\infty(g, b) \leq z \right\} \\ &= P \{ g F_\infty(g, b) \leq g z [1 + b_1 b_2 b_3 (c_1 + (g - 1) c_2)] \} + o(b_1 b_2 b_3) \\ &= G_g(g z [1 + b_1 b_2 b_3 (c_1 + (g - 1) c_2)]) \\ &+ A(g z [1 + b_1 b_2 b_3 (c_1 + (g - 1) c_2)]) b_1 b_2 b_3 + o(b_1 b_2 b_3) \\ &= G_g(g z) + G'_g(g z) g z [c_1 + (g - 1) c_2] b_1 b_2 b_3 + A(g z) b_1 b_2 b_3 + o(b_1 b_2 b_3) \\ &= G_g(g z) + G''_g(g z) g^2 z^2 c_2 b_1 b_2 b_3 + o(b_1 b_2 b_3). \end{aligned}$$

By definition,

$$\begin{aligned} P \{ F_{g, D-g+1} \leq z \} &= P \left\{ \chi_g^2 \leq g z \frac{\chi_{D-g+1}^2}{D - g + 1} \right\} = E G_g \left(g z \frac{\chi_{D-g+1}^2}{D - g + 1} \right) \\ &= G_g(g z) + G'_g(g z) g z E \left(\frac{\chi_{D-g+1}^2}{D - g + 1} - 1 \right) \\ &+ \frac{1}{2} G''_g(g z) \left(\frac{g z}{D - g + 1} \right)^2 E (\chi_{D-g+1}^2 - (D - g + 1))^2 + o \left(\frac{1}{D - g + 1} \right) \\ &= G_g(g z) + \frac{1}{D} G''_g(g z) g^2 z^2 + o \left(\frac{1}{D} \right) \\ &= G_g(g z) + G''_g(g z) g^2 z^2 c_2 b_1 b_2 b_3 + o(b_1 b_2 b_3) \end{aligned}$$

where we have used (A.13). Hence

$$P \left\{ \frac{\mu_1(D - g + 1)}{D} F_\infty(g, b) \leq z \right\} = P \{ F_{g, D-g+1} \leq z \} + o(b_1 b_2 b_3).$$

SUPPLEMENTARY APPENDIX

Sufficient Conditions for Assumptpion I5

In this section, we provide sufficient conditions for Assumption I5. Without loss of generality, we consider the case that J_{nT} is a scalar because $\sqrt{\frac{nT}{\ell_n \ell_T}} (\hat{J}_{nT} - \tilde{J}_{nT}) = o_p(1)$ holds if and only if $\sqrt{\frac{nT}{\ell_n \ell_T}} (a' \hat{J}_{nT} a - a' \tilde{J}_{nT} a) = o_p(1)$ for any $a \in \mathbb{R}^p$.

Assumption S1 (i) $\sqrt{nT} (\hat{\beta} - \beta_0) = O_p(1)$. (ii) $\sup_{i,t} E \left\| \left(\tilde{Z}_{it} \tilde{X}'_{it} \right) \right\|^2 < \infty$.
 (iii) $\sup_{j,s,\tilde{i},\tilde{t},\tilde{j},\tilde{s}} \sum_{i,t=1}^{\infty} \left\| E \left[\left(\tilde{Z}_{it}^{(c)} \tilde{u}_{it} \right) \left(\tilde{Z}_{js}^{(c)} \tilde{u}_{js} \right) \left(\tilde{Z}_{\tilde{i}\tilde{t}}^{(d)} \tilde{X}'_{\tilde{i}\tilde{t}} \right) \left(\tilde{Z}_{\tilde{j}\tilde{s}}^{(d)} \tilde{X}_{\tilde{j}\tilde{s}} \right) \right] \right\| < \infty$, for $c, d = 1, \dots, p$.

Assumption S2 Let

$$\mathfrak{U}_{(it,0)} = \tilde{Z}_{it} u_{it}, \quad \mathfrak{U}_{(it,1)} = \tilde{Z}_{it} \sqrt{T} \tilde{u}_i, \quad \mathfrak{U}_{(it,2)} = \tilde{Z}_{it} \sqrt{n} \tilde{u}_t.$$

For $\nu = 0, 1, 2$, $\varsigma = 1, 2$, c , and $d = 1, \dots, p$,

$$\begin{aligned} (i) & \sup_{\tilde{i}, \tilde{t}, \tilde{j}, \tilde{s}} \sum_{i,t=1}^{\infty} \sum_{j,s=1}^{\infty} \left\| E \left[\mathfrak{U}_{(it,\nu)}^{(c)} \mathfrak{U}_{(js,\varsigma)}^{(d)} \mathfrak{U}_{(\tilde{i}\tilde{t},\nu)}^{(c)} \mathfrak{U}_{(\tilde{j}\tilde{s},\varsigma)}^{(d)} \right] \right\| < \infty, \\ (ii) & \sup_{j,s,\tilde{i},\tilde{t},\tilde{j},\tilde{s}} \sum_{i,t=1}^{\infty} \left\| E \left[\mathfrak{U}_{(it,\nu)}^{(c)} \tilde{Z}_{js}^{(d)} \mathfrak{U}_{(\tilde{i}\tilde{t},\nu)}^{(c)} \tilde{Z}_{\tilde{j}\tilde{s}}^{(d)} \right] \right\| < \infty, \\ (iii) & \sup_{i,t,j,s,\tilde{i},\tilde{t},\tilde{j},\tilde{s}} \left\| E \left[\tilde{Z}_{it}^{(c)} \tilde{Z}_{js}^{(d)} \tilde{Z}_{\tilde{i}\tilde{t}}^{(c)} \tilde{Z}_{\tilde{j}\tilde{s}}^{(d)} \right] \right\| < \infty. \end{aligned}$$

Assumption S3 Let

$$\mathfrak{V}_{(it,0)} = V_{(i,t)}, \quad \mathfrak{V}_{(it,1)} = \sqrt{T} (\bar{Z}_i - \mu_{zi}) u_{it}, \quad \mathfrak{V}_{(it,2)} = \sqrt{n} (\bar{Z}_t - \mu_{zt}) u_{it}.$$

For $\nu = 0, 1, 2$, $\varsigma = 1, 2$, and $c, d = 1, \dots, p$,

$$\begin{aligned} (i) & \sup_{\tilde{i}, \tilde{t}, \tilde{j}, \tilde{s}} \sum_{i,t=1}^{\infty} \sum_{j,s=1}^{\infty} \left\| E \left[\mathfrak{V}_{(it,\nu)}^{(c)} \mathfrak{V}_{(js,\varsigma)}^{(d)} \mathfrak{V}_{(\tilde{i}\tilde{t},\nu)}^{(c)} \mathfrak{V}_{(\tilde{j}\tilde{s},\varsigma)}^{(d)} \right] \right\| < \infty, \\ (ii) & \sup_{j,s,\tilde{i},\tilde{t},\tilde{j},\tilde{s}} \sum_{i,t=1}^{\infty} \left\| E \left[\mathfrak{V}_{(it,\nu)}^{(c)} \mathfrak{V}_{(\tilde{i}\tilde{t},\nu)}^{(c)} u_{js} u_{\tilde{j}\tilde{s}} \right] \right\| < \infty, \\ (iii) & \sup_{i,t,j,s,\tilde{i},\tilde{t},\tilde{j},\tilde{s}} \left\| E \left[u_{it} u_{js} u_{\tilde{i}\tilde{t}} u_{\tilde{j}\tilde{s}} \right] \right\| < \infty. \end{aligned}$$

Lemma S1 Under Assumptions I4, S1, S2, and S3,

$$\sqrt{\frac{nT}{\ell_n \ell_T}} (\hat{J}_{nT} - \tilde{J}_{nT}) = o_p(1).$$

Proof. As

$$\sqrt{\frac{nT}{\ell_n \ell_T}} (\hat{J}_{nT} - \tilde{J}_{nT}) = \sqrt{\frac{nT}{\ell_n \ell_T}} (\hat{J}_{nT} - \tilde{J}_{nT}^*) + \sqrt{\frac{nT}{\ell_n \ell_T}} (\tilde{J}_{nT}^* - \tilde{J}_{nT}^o) + \sqrt{\frac{nT}{\ell_n \ell_T}} (\tilde{J}_{nT}^o - \tilde{J}_{nT})$$

where

$$\begin{aligned} \tilde{J}_{nT}^* &= \frac{1}{nT} \sum_{i,j=1}^n \sum_{t,s=1}^T K_1 \left(\frac{d_{ij}}{d_n} \right) K_2 \left(\frac{d_{ts}}{d_T} \right) \tilde{Z}_{it} \tilde{u}_{it} \tilde{Z}'_{js} \tilde{u}_{js} \\ \tilde{J}_{nT}^o &= \frac{1}{nT} \sum_{i,j=1}^n \sum_{t,s=1}^T K_1 \left(\frac{d_{ij}}{d_n} \right) K_2 \left(\frac{d_{ts}}{d_T} \right) \tilde{Z}_{it} u_{it} \tilde{Z}'_{js} u_{js}, \end{aligned}$$

it is sufficient to show that

$$\sqrt{\frac{nT}{\ell_n \ell_T}} \left(\hat{J}_{nT} - \tilde{J}_{nT}^* \right) = o_p(1), \quad (\text{S.1})$$

$$\sqrt{\frac{nT}{\ell_n \ell_T}} \left(\tilde{J}_{nT}^* - \tilde{J}_{nT}^o \right) = o_p(1), \quad (\text{S.2})$$

$$\sqrt{\frac{nT}{\ell_n \ell_T}} \left(\tilde{J}_{nT}^o - \tilde{J}_{nT} \right) = o_p(1). \quad (\text{S.3})$$

First, in order to show (S.1), we define

$$J_{nT}(\beta) = \frac{1}{nT} \sum_{i,j=1}^n \sum_{t,s=1}^T K_1 \left(\frac{d_{ij}}{d_n} \right) K_2 \left(\frac{d_{ts}}{d_T} \right) \tilde{V}_{(i,t)}(\beta) \tilde{V}_{(j,s)}(\beta), \text{ with } \tilde{V}_{(i,t)}(\beta) = \tilde{Z}_{it} \left(\tilde{Y}_{it} - \tilde{X}_{it}' \beta \right).$$

This implies

$$\tilde{J}_{nT}^* = J_{nT}(\beta_0), \quad \tilde{V}_{(i,t)}(\beta_0) = \tilde{Z}_{it} \tilde{u}_{it}, \text{ and } \frac{\partial}{\partial \beta'} \tilde{V}_{(i,t)}(\beta) = -\tilde{Z}_{it} \tilde{X}_{it}'.$$

Using the Taylor expansion, we obtain

$$\begin{aligned} \sqrt{\frac{nT}{\ell_n \ell_T}} \left(\hat{J}_{nT} - \tilde{J}_{nT}^* \right) &= \left[\frac{\partial}{\partial \beta'} J_{nT}(\beta_0) / \sqrt{\ell_n \ell_T} \right] \sqrt{nT} (\hat{\beta} - \beta_0) \\ &\quad + \frac{1}{2} \left(\sqrt{nT} (\hat{\beta} - \beta_0) \right)' \left[\frac{\partial^2}{\partial \beta \partial \beta'} J_{nT}(\bar{\beta}) / \sqrt{\ell_n \ell_T nT} \right] \sqrt{nT} (\hat{\beta} - \beta_0) \\ &:= H'_{1nT} \sqrt{nT} (\hat{\beta} - \beta_0) + \frac{1}{2} \left(\sqrt{nT} (\hat{\beta} - \beta_0) \right)' H_{2nT} \sqrt{nT} (\hat{\beta} - \beta_0). \end{aligned} \quad (\text{S.4})$$

where $\bar{\beta}$ lies on the line segment joining $\hat{\beta}$ and β_0 . For H_{1nT} ,

$$\begin{aligned} H'_{1nT} &= \frac{1}{\sqrt{\ell_n \ell_T}} \frac{1}{nT} \sum_{i,j=1}^n \sum_{t,s=1}^T K_1 \left(\frac{d_{ij}}{d_n} \right) K_2 \left(\frac{d_{ts}}{d_T} \right) \frac{\partial}{\partial \beta'} \left(\tilde{V}_{(i,t)}(\beta_0) \tilde{V}_{(j,s)}(\beta_0) \right) \\ &= \frac{2}{\sqrt{\ell_n \ell_T}} \frac{1}{nT} \sum_{i,j=1}^n \sum_{t,s=1}^T K_1 \left(\frac{d_{ij}}{d_n} \right) K_2 \left(\frac{d_{ts}}{d_T} \right) \tilde{V}_{(i,t)}(\beta_0) \frac{\partial}{\partial \beta'} \tilde{V}_{(j,s)}(\beta_0) \\ &= -2 \sqrt{\frac{\ell_n \ell_T}{nT}} \frac{1}{\ell_n \ell_T \sqrt{nT}} \sum_{i,j=1}^n \sum_{t,s=1}^T K_1 \left(\frac{d_{ij}}{d_n} \right) K_2 \left(\frac{d_{ts}}{d_T} \right) \left(\tilde{Z}_{it} \tilde{u}_{it} \right) \left(\tilde{Z}_{js} \tilde{X}_{js}' \right). \end{aligned}$$

By Markov inequality and Assumptions I4 and S1(iii),

$$\begin{aligned}
& P \left(\left\| \frac{1}{\ell_n \ell_T \sqrt{nT}} \sum_{i,j=1}^n \sum_{t,s=1}^T K_1 \left(\frac{d_{ij}}{d_n} \right) K_2 \left(\frac{d_{ts}}{d_T} \right) (\tilde{Z}_{it} \tilde{u}_{it}) (\tilde{Z}_{js} \tilde{X}'_{js}) \right\| > \Delta \right) \\
& \leq \frac{1}{\Delta^2} E \left[\frac{1}{\ell_n^2 \ell_T^2 nT} \sum_{i,j,t,s} K_1 \left(\frac{d_{ij}}{d_n} \right) K_2 \left(\frac{d_{ts}}{d_T} \right) K_1 \left(\frac{d_{i\tilde{j}}}{d_n} \right) K_2 \left(\frac{d_{t\tilde{s}}}{d_T} \right) (\tilde{Z}_{it} \tilde{u}_{it}) (\tilde{Z}_{i\tilde{t}} \tilde{u}_{i\tilde{t}}) \tilde{Z}_{js} \tilde{X}'_{js} (\tilde{Z}_{j\tilde{s}} \tilde{X}'_{j\tilde{s}})' \right] \\
& \leq \frac{1}{\Delta^2} \frac{\ell_{i,n} \ell_{t,T} \ell_{i,n} \ell_{t,T}}{\ell_n^2 \ell_T^2} \frac{1}{\ell_{i,n} \ell_{t,T} \ell_{i,n} \ell_{t,T} nT} \sum_{i,\tilde{i}=1}^n \sum_{t,\tilde{t}=1}^T \sum_{j(i)=1}^{\ell_{i,n}} \sum_{s(t)=1}^{\ell_{t,T}} \sum_{\tilde{j}(\tilde{i})=1}^{\ell_{i,n}} \sum_{\tilde{s}(\tilde{t})=1}^{\ell_{t,T}} \left\| E \left[(\tilde{Z}_{it} \tilde{u}_{it}) (\tilde{Z}_{i\tilde{t}} \tilde{u}_{i\tilde{t}}) \tilde{Z}_{js} \tilde{X}'_{js} (\tilde{Z}_{j\tilde{s}} \tilde{X}'_{j\tilde{s}})' \right] \right\| \\
& \leq \frac{1}{\Delta^2} \frac{\ell_{i,n} \ell_{t,T} \ell_{i,n} \ell_{t,T}}{\ell_n^2 \ell_T^2} \sup_{j,s,\tilde{i},\tilde{t},\tilde{j},\tilde{s}} \sum_{i,t=1}^{\infty} \left\| E \left[(\tilde{Z}_{it} \tilde{u}_{it}) (\tilde{Z}_{i\tilde{t}} \tilde{u}_{i\tilde{t}}) (\tilde{Z}_{js} \tilde{X}'_{js}) (\tilde{Z}_{j\tilde{s}} \tilde{X}'_{j\tilde{s}})' \right] \right\| \\
& \leq \frac{C}{\Delta^2},
\end{aligned}$$

for some C and thus we have

$$H_{1nT} = O_p \left(\sqrt{\ell_n \ell_T / nT} \right) = o_p(1). \quad (\text{S.5})$$

For H_{2nT} ,

$$\begin{aligned}
H_{2nT} &= \frac{2}{\sqrt{\ell_n \ell_T nT}} \frac{1}{nT} \sum_{i,j=1}^n \sum_{t,s=1}^T K_1 \left(\frac{d_{ij}}{d_n} \right) K_2 \left(\frac{d_{ts}}{d_T} \right) \left(\frac{\partial}{\partial \beta} \tilde{V}_{(i,t)}(\bar{\beta}) \right) \left(\frac{\partial}{\partial \beta'} \tilde{V}_{(j,s)}(\bar{\beta}) \right) \\
&= \frac{2}{\sqrt{\ell_n \ell_T nT}} \frac{1}{nT} \sum_{i,j=1}^n \sum_{t,s=1}^T K_1 \left(\frac{d_{ij}}{d_n} \right) K_2 \left(\frac{d_{ts}}{d_T} \right) (\tilde{Z}_{it} \tilde{X}'_{it})' (\tilde{Z}_{js} \tilde{X}'_{js}) \\
&= o_p(1), \quad (\text{S.6})
\end{aligned}$$

because

$$\begin{aligned}
\|H_{2nT}\|^2 &= \left\| \frac{2}{\sqrt{\ell_n \ell_T nT}} \frac{1}{nT} \sum_{i,j=1}^n \sum_{t,s=1}^T K_1 \left(\frac{d_{ij}}{d_n} \right) K_2 \left(\frac{d_{ts}}{d_T} \right) (\tilde{Z}_{it} \tilde{X}'_{it})' (\tilde{Z}_{js} \tilde{X}'_{js}) \right\|^2 \\
&\leq \left[\frac{2}{\sqrt{\ell_n \ell_T nT}} \frac{1}{nT} \sum_{i,j=1}^n \sum_{t,s=1}^T K_1 \left(\frac{d_{ij}}{d_n} \right) K_2 \left(\frac{d_{ts}}{d_T} \right) \left\| (\tilde{Z}_{it} \tilde{X}'_{it})' \right\| \left\| \tilde{Z}_{js} \tilde{X}'_{js} \right\| \right]^2 \\
&\leq \left[2 \sqrt{\frac{\ell_n \ell_T}{nT}} \frac{1}{nT} \sum_{i,t} \left\| (\tilde{Z}_{it} \tilde{X}'_{it})' \right\| \frac{1}{\ell_n \ell_T} \sum_{j,s} 1 \{d_{ij} \leq d_n\} 1 \{d_{ts} \leq d_T\} \left\| \tilde{Z}_{js} \tilde{X}'_{js} \right\| \right]^2 \\
&\leq 4 \frac{\ell_n \ell_T}{nT} \left(\frac{1}{nT} \sum_{i,t} \left\| (\tilde{Z}_{it} \tilde{X}'_{it})' \right\|^2 \right) \frac{1}{nT} \sum_{i,t} \left(\frac{1}{\ell_n \ell_T} \sum_{j,s} 1 \{d_{ij} \leq d_n\} 1 \{d_{ts} \leq d_T\} \left\| \tilde{Z}_{js} \tilde{X}'_{js} \right\|^2 \right) \\
&= O_p \left(\frac{\ell_n \ell_T}{nT} \right).
\end{aligned}$$

by Assumption S1(ii). From (S.4), (S.5), (S.6) and Assumption S1(i), we have

$$\sqrt{\frac{nT}{\ell_n \ell_T}} \left(\hat{J}_{nT} - \tilde{J}_{nT}^* \right) = o_p(1).$$

Secondly, we show (S.2).

$$\begin{aligned} & \sqrt{\frac{nT}{\ell_n \ell_T}} \left(\tilde{J}_{nT}^* - \tilde{J}_{nT}^o \right) \\ &= \sqrt{\frac{nT}{\ell_n \ell_T}} \frac{1}{nT} \sum_{i,j=1}^n \sum_{t,s=1}^T K_1 \left(\frac{d_{ij}}{d_n} \right) K_2 \left(\frac{d_{ts}}{d_T} \right) \tilde{Z}_{it} \tilde{Z}_{js} (\tilde{u}_{it} \tilde{u}_{js} - u_{it} u_{js}) \\ &= \sqrt{\frac{nT}{\ell_n \ell_T}} \frac{1}{nT} \sum_{i,j=1}^n \sum_{t,s=1}^T K_1 \left(\frac{d_{ij}}{d_n} \right) K_2 \left(\frac{d_{ts}}{d_T} \right) \tilde{Z}_{it} \tilde{Z}_{js} [-2(u_{it} \bar{u}_j + u_{it} \bar{u}_s - u_{it} \bar{u}) + (\bar{u}_i + \bar{u}_t - \bar{u})(\bar{u}_j + \bar{u}_s - \bar{u})] \\ &= -2 \sqrt{\frac{nT}{\ell_n \ell_T}} \frac{1}{nT} \sum_{i,j=1}^n \sum_{t,s=1}^T K_1 \left(\frac{d_{ij}}{d_n} \right) K_2 \left(\frac{d_{ts}}{d_T} \right) \left(\frac{1}{\sqrt{T}} \mathfrak{U}_{(it,0)} \mathfrak{U}_{(js,1)} + \frac{1}{\sqrt{n}} \mathfrak{U}_{(it,0)} \mathfrak{U}_{(js,2)} - \mathfrak{U}_{(it,0)} \tilde{Z}_{js} \bar{u} \right) \\ &+ \sqrt{\frac{nT}{\ell_n \ell_T}} \frac{1}{nT} \sum_{i,j=1}^n \sum_{t,s=1}^T K_1 \left(\frac{d_{ij}}{d_n} \right) K_2 \left(\frac{d_{ts}}{d_T} \right) \left[\frac{1}{n} \mathfrak{U}_{(it,1)} \mathfrak{U}_{(js,1)} + \frac{1}{T} \mathfrak{U}_{(it,2)} \mathfrak{U}_{(js,2)} + \left(\tilde{Z}_{it} \bar{u} \right) \left(\tilde{Z}_{js} \bar{u} \right) \right. \\ &\left. + \frac{2}{\sqrt{nT}} \mathfrak{U}_{(it,1)} \mathfrak{U}_{(js,2)} - \frac{2}{\sqrt{T}} \mathfrak{U}_{(it,1)} \left(\tilde{Z}_{js} \bar{u} \right) - \frac{2}{\sqrt{n}} \mathfrak{U}_{(it,2)} \left(\tilde{Z}_{js} \bar{u} \right) \right]. \end{aligned} \quad (\text{S.7})$$

We have

$$\sqrt{\frac{nT}{\ell_n \ell_T}} \frac{1}{nT} \sum_{i,j=1}^n \sum_{t,s=1}^T K_1 \left(\frac{d_{ij}}{d_n} \right) K_2 \left(\frac{d_{ts}}{d_T} \right) \frac{1}{\sqrt{T}} \mathfrak{U}_{(it,0)} \mathfrak{U}_{(js,1)} = o_p(1),$$

because

$$\begin{aligned} & P \left(\left\| \sqrt{\frac{nT}{\ell_n \ell_T}} \frac{1}{nT} \sum_{i,j=1}^n \sum_{t,s=1}^T K_1 \left(\frac{d_{ij}}{d_n} \right) K_2 \left(\frac{d_{ts}}{d_T} \right) \frac{1}{\sqrt{T}} \mathfrak{U}_{(it,0)} \mathfrak{U}_{(js,1)} \right\| > \Delta \right) \\ & \leq \frac{1}{T} \frac{1}{\Delta^2} \frac{1}{\ell_n \ell_T nT} \sum_{i,\tilde{i}=1}^n \sum_{t,\tilde{t}=1}^T \sum_{j(i)=1}^{\ell_{i,n}} \sum_{s(t)=1}^{\ell_{t,T}} \sum_{\tilde{j}(\tilde{i})=1}^{\ell_{\tilde{i},n}} \sum_{\tilde{s}(\tilde{t})=1}^{\ell_{\tilde{t},T}} \left\| E \left[\mathfrak{U}_{(it,0)} \mathfrak{U}_{(j(i)s(t),1)} \mathfrak{U}_{(\tilde{i}\tilde{t},0)} \mathfrak{U}_{(\tilde{j}(\tilde{i})\tilde{s}(\tilde{t}),1)} \right] \right\| \\ & \leq \frac{1}{T} \frac{C}{\Delta^2} \sup_{\tilde{i},\tilde{t},\tilde{j},\tilde{s}} \sum_{i,t=1}^{\infty} \sum_{j,s=1}^{\infty} \left\| E \left[\mathfrak{U}_{(it,0)} \mathfrak{U}_{(js,1)} \mathfrak{U}_{(\tilde{i}\tilde{t},0)} \mathfrak{U}_{(\tilde{j}\tilde{s},1)} \right] \right\| \\ & = o(1) \end{aligned}$$

by Assumptions I4 and S2(i). Second,

$$\begin{aligned} & \sqrt{\frac{nT}{\ell_n \ell_T}} \frac{1}{nT} \sum_{i,j=1}^n \sum_{t,s=1}^T K_1 \left(\frac{d_{ij}}{d_n} \right) K_2 \left(\frac{d_{ts}}{d_T} \right) \mathfrak{U}_{(it,0)} \left(\tilde{Z}_{js} \bar{u} \right) \\ &= \left(\sqrt{nT} \bar{u} \right) \sqrt{\frac{\ell_n \ell_T}{nT}} \frac{1}{\ell_n \ell_T \sqrt{nT}} \sum_{i,j=1}^n \sum_{t,s=1}^T K_1 \left(\frac{d_{ij}}{d_n} \right) K_2 \left(\frac{d_{ts}}{d_T} \right) \mathfrak{U}_{(it,0)} \left(\tilde{Z}_{js} \right) \\ &= o_p(1) \end{aligned}$$

because

$$\begin{aligned}
& P \left(\left\| \frac{1}{\ell_n \ell_T \sqrt{nT}} \sum_{i,j=1}^n \sum_{t,s=1}^T K_1 \left(\frac{d_{ij}}{d_n} \right) K_2 \left(\frac{d_{ts}}{d_T} \right) \mathfrak{U}_{(it,0)} \tilde{Z}_{js} \right\| > \Delta \right) \\
& \leq \frac{1}{\Delta^2} \frac{1}{(\ell_n \ell_T)^2 nT} \sum_{i,\tilde{i}=1}^n \sum_{t,\tilde{t}=1}^T \sum_{j(i)=1}^{\ell_{i,n}} \sum_{s(t)=1}^{\ell_{t,T}} \sum_{\tilde{j}(\tilde{i})=1}^{\ell_{\tilde{i},n}} \sum_{\tilde{s}(\tilde{t})=1}^{\ell_{\tilde{t},T}} \left\| E \left[\mathfrak{U}_{(it,0)} \tilde{Z}_{j(i)s(t)} \mathfrak{U}_{(\tilde{i}\tilde{t},0)} \tilde{Z}_{\tilde{j}(\tilde{i})\tilde{s}(\tilde{t})} \right] \right\| \\
& \leq \frac{C}{\Delta^2}
\end{aligned}$$

by Assumptions I4 and S2(ii). Using the same procedures, we can show that the other terms in (S.7) are $o_p(1)$. Thus,

$$\sqrt{\frac{nT}{\ell_n \ell_T}} \left(\tilde{J}_{nT}^* - \tilde{J}_{nT}^o \right) = o_p(1).$$

Finally we show (S.3). We have

$$\begin{aligned}
& \sqrt{\frac{nT}{\ell_n \ell_T}} \left(\tilde{J}_{nT}^o - \tilde{J}_{nT} \right) \\
& = \sqrt{\frac{nT}{\ell_n \ell_T}} \frac{1}{nT} \sum_{i,j=1}^n \sum_{t,s=1}^T K_1 \left(\frac{d_{ij}}{d_n} \right) K_2 \left(\frac{d_{ts}}{d_T} \right) [(Z_{it} - \bar{Z}_i - \bar{Z}_t + \bar{Z}) (Z_{js} - \bar{Z}_j - \bar{Z}_s + \bar{Z}) u_{it} u_{js} \\
& \quad - (Z_{it} - \mu_{zi} - \mu_{zt} + \mu_z) (Z_{js} - \mu_{zj} - \mu_{zs} + \mu_z) u_{it} u_{js}] \\
& = -2 \sqrt{\frac{nT}{\ell_n \ell_T}} \frac{1}{nT} \sum_{i,j=1}^n \sum_{t,s=1}^T K_1 \left(\frac{d_{ij}}{d_n} \right) K_2 \left(\frac{d_{ts}}{d_T} \right) \left[\frac{1}{\sqrt{T}} V_{(i,t)} \mathfrak{V}_{(js,1)} + \frac{1}{\sqrt{n}} V_{(i,t)} \mathfrak{V}_{(js,2)} - V_{(i,t)} (\bar{Z} - \mu_z) u_{js} \right] \\
& \quad + \sqrt{\frac{nT}{\ell_n \ell_T}} \frac{1}{nT} \sum_{i,j=1}^n \sum_{t,s=1}^T K_1 \left(\frac{d_{ij}}{d_n} \right) K_2 \left(\frac{d_{ts}}{d_T} \right) \left[\frac{1}{T} \mathfrak{V}_{(it,1)} \mathfrak{V}_{(js,1)} + \frac{1}{n} \mathfrak{V}_{(it,2)} \mathfrak{V}_{(js,2)} + (\bar{Z} - \mu_z) u_{it} (\bar{Z} - \mu_z) u_{js} \right. \\
& \quad \left. + \frac{2}{\sqrt{nT}} \mathfrak{V}_{(it,1)} \mathfrak{V}_{(js,2)} - \frac{2}{\sqrt{T}} \mathfrak{V}_{(it,1)} (\bar{Z} - \mu_z) u_{js} - \frac{2}{\sqrt{n}} \mathfrak{V}_{(it,2)} (\bar{Z} - \mu_z) u_{js} \right]. \tag{S.8}
\end{aligned}$$

First, we have

$$\frac{1}{\sqrt{T}} \sqrt{\frac{nT}{\ell_n \ell_T}} \frac{1}{nT} \sum_{i,j=1}^n \sum_{t,s=1}^T K_1 \left(\frac{d_{ij}}{d_n} \right) K_2 \left(\frac{d_{ts}}{d_T} \right) V_{(i,t)} \mathfrak{V}_{(js,1)} = o_p(1),$$

because

$$\begin{aligned}
& P \left(\left\| \frac{1}{\sqrt{T}} \sqrt{\frac{nT}{\ell_n \ell_T}} \frac{1}{nT} \sum_{i,j=1}^n \sum_{t,s=1}^T K_1 \left(\frac{d_{ij}}{d_n} \right) K_2 \left(\frac{d_{ts}}{d_T} \right) V_{(i,t)} \mathfrak{V}_{(js,1)} \right\| > \Delta \right) \\
& \leq \frac{1}{T} \frac{1}{\Delta^2} \frac{1}{\ell_n \ell_T nT} \sum_{i,\tilde{i}=1}^n \sum_{t,\tilde{t}=1}^T \sum_{j(i)=1}^{\ell_{i,n}} \sum_{s(t)=1}^{\ell_{t,T}} \sum_{\tilde{j}(\tilde{i})=1}^{\ell_{\tilde{i},n}} \sum_{\tilde{s}(\tilde{t})=1}^{\ell_{\tilde{t},T}} \left\| E \left[V_{(i,t)} \mathfrak{V}_{(j(i)s(t),1)} V_{(\tilde{i},\tilde{t})} \mathfrak{V}_{(\tilde{j}(\tilde{i})\tilde{s}(\tilde{t}),1)} \right] \right\| \\
& \leq \frac{1}{T} \frac{C}{\Delta^2} \sup_{\tilde{i},\tilde{t},\tilde{j},\tilde{s}} \sum_{i,t=1}^{\infty} \sum_{j,s=1}^{\infty} \left\| E \left[V_{(i,t)} \mathfrak{V}_{(js,1)} V_{(\tilde{i},\tilde{t})} \mathfrak{V}_{(\tilde{j}\tilde{s},1)} \right] \right\| \\
& = o(1)
\end{aligned}$$

by Assumptions I4 and S3(i). Using the similar procedures, we can show that the other terms in (S.8) are also $o_p(1)$ under Assumptions I4 and S3. This implies

$$\sqrt{\frac{nT}{\ell_n \ell_T}} \left(\tilde{J}_{nT}^o - \tilde{J}_{nT} \right) = o_p(1)$$

and the proof of Lemma S1 is complete. ■

Proof of Lemma 1

We can decompose $\mathfrak{X}_{\iota, nT}$ as follows.

$$\begin{aligned} \mathfrak{X}_{\iota, nT} &= \frac{1}{\sqrt{nT}} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^T \Phi_{b,\iota} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) 1_{i_1, i_2} \tilde{Z}_{(i_1, i_2, t)} \left(u_{(i_1, i_2, t)} - \bar{u}_{(i_1, i_2)} - \bar{u}_t + \bar{u} \right) \\ &\quad - \left(\frac{1}{nT} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^T \Phi_{b,\iota} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) 1_{i_1, i_2} \tilde{Z}_{(i_1, i_2, t)} \tilde{X}'_{(i_1, i_2, t)} \right) \\ &\quad \times \left(\frac{1}{nT} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^T 1_{i_1, i_2} \tilde{Z}_{(i_1, i_2, t)} \tilde{X}'_{(i_1, i_2, t)} \right)^{-1} \left(\frac{1}{\sqrt{nT}} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^T 1_{i_1, i_2} \tilde{Z}_{(i_1, i_2, t)} u_{(i_1, i_2, t)} \right). \end{aligned} \quad (\text{S.9})$$

The first term in the rhs of (S.9) is equal to

$$\begin{aligned} &\frac{1}{\sqrt{nT}} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^T \Phi_{b,\iota} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) 1_{i_1, i_2} \tilde{Z}_{(i_1, i_2, t)} \left(u_{(i_1, i_2, t)} - \bar{u}_{(i_1, i_2)} - \bar{u}_t + \bar{u} \right) \\ &\quad - \frac{1}{\sqrt{nT}} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^T \Phi_{b,\iota} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) 1_{i_1, i_2} \left[(\bar{Z}_{(i_1, i_2)} - \mu_{z(i_1, i_2)}) + (\bar{Z}_t - \mu_{zt}) - (\bar{Z} - \mu_z) \right] \\ &\quad \times \left(u_{(i_1, i_2, t)} - \bar{u}_{(i_1, i_2)} - \bar{u}_t + \bar{u} \right) \\ &\quad := A_1 + A_2. \end{aligned}$$

Among the terms in A_1 , since

$$\begin{aligned} &E \left[\frac{1}{\sqrt{nT}} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^T \Phi_{b,\iota} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) 1_{i_1, i_2} \tilde{Z}_{(i_1, i_2, t)} \bar{u}_{(i_1, i_2)} \right]^2 \\ &\leq \frac{1}{T} \frac{C}{nT} \sum_{i,t} \sum_{j,s} \left\| E \tilde{Z}_{it} \sqrt{T} \bar{u}_i \tilde{Z}_{js} \sqrt{T} \bar{u}_j \right\| \\ &= o(1), \end{aligned}$$

we have

$$\frac{1}{\sqrt{nT}} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^T \Phi_{b,\iota} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) 1_{i_1, i_2} \tilde{Z}_{(i_1, i_2, t)} \bar{u}_{(i_1, i_2)} = o_p(1),$$

for $\iota = 1, \dots, \mathcal{L}$. Using the similar procedure, we can also show that

$$\frac{1}{\sqrt{nT}} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^T \Phi_{b,\iota} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) 1_{i_1, i_2} \tilde{Z}_{(i_1, i_2, t)} (\bar{u}_t - \bar{u}) = o_p(1).$$

Combining the above two results yields:

$$A_1 = \frac{1}{\sqrt{nT}} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^T \Phi_{b,\ell} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) V_{(i_1, i_2, t)}^* + o_p(1).$$

Similarly, it is straightforward to show $A_2 = o_p(1)$. As a result, $\mathfrak{X}_{\ell, nT}$ can be rewritten as

$$\begin{aligned} \mathfrak{X}_{\ell, nT} &= \frac{1}{\sqrt{nT}} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^T \Phi_{b,\ell} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) V_{(i_1, i_2, t)}^* \\ &\quad - \left(\frac{1}{nT} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^T \Phi_{b,\ell} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) 1_{i_1, i_2} \tilde{Z}_{(i_1, i_2, t)} \tilde{X}'_{(i_1, i_2, t)} \right) \\ &\quad \times \left(\frac{1}{nT} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^T 1_{i_1, i_2} \tilde{Z}_{(i_1, i_2, t)} \tilde{X}'_{(i_1, i_2, t)} \right)^{-1} \frac{1}{\sqrt{nT}} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^T 1_{i_1, i_2} \tilde{Z}_{(i_1, i_2, t)} u_{(i_1, i_2, t)} + o_p(1). \end{aligned} \quad (\text{S.10})$$

Next,

$$\begin{aligned} &\frac{1}{nT} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^T \Phi_{b,\ell} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) 1_{i_1, i_2} \tilde{Z}_{(i_1, i_2, t)} \tilde{X}'_{(i_1, i_2, t)} \\ &= \frac{1}{nT} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^T \Phi_{b,\ell} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) \left[\sum_{a_1=1}^{i_1} \sum_{a_2=1}^{i_2} \sum_{a_3=1}^t 1_{i_1, i_2} \tilde{Z}_{(i_1, i_2, t)} \tilde{X}'_{(i_1, i_2, t)} - \sum_{a_1=1}^{i_1-1} \sum_{a_2=1}^{i_2} \sum_{a_3=1}^t 1_{i_1, i_2} \tilde{Z}_{(i_1, i_2, t)} \tilde{X}'_{(i_1, i_2, t)} \right] \\ &\quad - \frac{1}{nT} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^T \Phi_{b,\ell} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) \left[\sum_{a_1=1}^{i_1} \sum_{a_2=1}^{i_2-1} \sum_{a_3=1}^t 1_{i_1, i_2} \tilde{Z}_{(i_1, i_2, t)} \tilde{X}'_{(i_1, i_2, t)} - \sum_{a_1=1}^{i_1-1} \sum_{a_2=1}^{i_2-1} \sum_{a_3=1}^t 1_{i_1, i_2} \tilde{Z}_{(i_1, i_2, t)} \tilde{X}'_{(i_1, i_2, t)} \right] \\ &\quad - \frac{1}{nT} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^T \Phi_{b,\ell} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) \left[\sum_{a_1=1}^{i_1} \sum_{a_2=1}^{i_2} \sum_{a_3=1}^{t-1} 1_{i_1, i_2} \tilde{Z}_{(i_1, i_2, t)} \tilde{X}'_{(i_1, i_2, t)} - \sum_{a_1=1}^{i_1} \sum_{a_2=1}^{i_2-1} \sum_{a_3=1}^{t-1} 1_{i_1, i_2} \tilde{Z}_{(i_1, i_2, t)} \tilde{X}'_{(i_1, i_2, t)} \right] \\ &\quad + \frac{1}{nT} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^T \Phi_{b,\ell} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) \left[\sum_{a_1=1}^{i_1-1} \sum_{a_2=1}^{i_2} \sum_{a_3=1}^{t-1} 1_{i_1, i_2} \tilde{Z}_{(i_1, i_2, t)} \tilde{X}'_{(i_1, i_2, t)} - \sum_{a_1=1}^{i_1-1} \sum_{a_2=1}^{i_2-1} \sum_{a_3=1}^{t-1} 1_{i_1, i_2} \tilde{Z}_{(i_1, i_2, t)} \tilde{X}'_{(i_1, i_2, t)} \right] \\ &:= I_1 + I_2 + I_3 + I_4 \end{aligned}$$

For I_1 , we have

$$\begin{aligned} I_1 &= \frac{1}{nT} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^T \left[\Phi_{b,\ell} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) - \Phi_{b,\ell} \left(\frac{i_1+1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) \right] \sum_{a_1=1}^{i_1} \sum_{a_2=1}^{i_2} \sum_{a_3=1}^t 1_{i_1, i_2} \tilde{Z}_{(i_1, i_2, t)} \tilde{X}'_{(i_1, i_2, t)} \\ &\quad + \frac{1}{nT} \sum_{i_2=1}^{M_n} \sum_{t=1}^T \Phi_{b,\ell} \left(\frac{L_n+1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) \sum_{a_1=1}^{L_n} \sum_{a_2=1}^{i_2} \sum_{a_3=1}^t 1_{i_1, i_2} \tilde{Z}_{(i_1, i_2, t)} \tilde{X}'_{(i_1, i_2, t)} \\ &= \frac{1}{nT} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^T \left[\Phi_{b,\ell} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) - \Phi_{b,\ell} \left(\frac{i_1+1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) \right] i_1 i_2 t Q (1 + o_p(1)) \\ &\quad + \frac{1}{nT} \sum_{i_2=1}^{M_n} \sum_{t=1}^T \Phi_{b,\ell} \left(\frac{L_n+1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) L_n i_2 t Q (1 + o_p(1)). \end{aligned} \quad (\text{S.11})$$

For I_2 , we have

$$\begin{aligned}
I_2 &= -\frac{1}{nT} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^T \left[\Phi_{b,\ell} \left(\frac{i_1}{L_n}, \frac{i_2+1}{M_n}, \frac{t}{T} \right) - \Phi_{b,\ell} \left(\frac{i_1+1}{L_n}, \frac{i_2+1}{M_n}, \frac{t}{T} \right) \right] \sum_{a_1=1}^{i_1} \sum_{a_2=1}^{i_2} \sum_{a_3=1}^t 1_{i_1,i_2} \tilde{Z}_{(i_1,i_2,t)} \tilde{X}'_{(i_1,i_2,t)} \\
&+ \frac{1}{nT} \sum_{i_1=1}^{L_n} \sum_{t=1}^T \left[\Phi_{b,\ell} \left(\frac{i_1}{L_n}, \frac{M_n+1}{M_n}, \frac{t}{T} \right) - \Phi_{b,\ell} \left(\frac{i_1+1}{L_n}, \frac{M_n+1}{M_n}, \frac{t}{T} \right) \right] \sum_{a_1=1}^{i_1} \sum_{a_2=1}^{M_n} \sum_{a_3=1}^t 1_{i_1,i_2} \tilde{Z}_{(i_1,i_2,t)} \tilde{X}'_{(i_1,i_2,t)} \\
&- \frac{1}{nT} \sum_{i_2=1}^{M_n} \sum_{t=1}^T \Phi_{b,\ell} \left(\frac{L_n+1}{L_n}, \frac{i_2+1}{M_n}, \frac{t}{T} \right) \sum_{a_1=1}^{L_n} \sum_{a_2=1}^{i_2} \sum_{a_3=1}^t 1_{i_1,i_2} \tilde{Z}_{(i_1,i_2,t)} \tilde{X}'_{(i_1,i_2,t)} \\
&+ \frac{1}{nT} \sum_{t=1}^T \Phi_{b,\ell} \left(\frac{L_n+1}{L_n}, \frac{M_n+1}{M_n}, \frac{t}{T} \right) \sum_{a_1=1}^{L_n} \sum_{a_2=1}^{M_n} \sum_{a_3=1}^t 1_{i_1,i_2} \tilde{Z}_{(i_1,i_2,t)} \tilde{X}'_{(i_1,i_2,t)} \\
&= -\frac{1}{nT} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^T \left[\Phi_{b,\ell} \left(\frac{i_1}{L_n}, \frac{i_2+1}{M_n}, \frac{t}{T} \right) - \Phi_{b,\ell} \left(\frac{i_1+1}{L_n}, \frac{i_2+1}{M_n}, \frac{t}{T} \right) \right] i_1 i_2 t Q (1 + o_p(1)) \\
&+ \frac{1}{nT} \sum_{i_1=1}^{L_n} \sum_{t=1}^T \left[\Phi_{b,\ell} \left(\frac{i_1}{L_n}, \frac{M_n+1}{M_n}, \frac{t}{T} \right) - \Phi_{b,\ell} \left(\frac{i_1+1}{L_n}, \frac{M_n+1}{M_n}, \frac{t}{T} \right) \right] i_1 M_n t Q (1 + o_p(1)) \\
&- \frac{1}{nT} \sum_{i_2=1}^{M_n} \sum_{t=1}^T \Phi_{b,\ell} \left(\frac{L_n+1}{L_n}, \frac{i_2+1}{M_n}, \frac{t}{T} \right) L_n i_2 t Q (1 + o_p(1)) \\
&+ \frac{1}{nT} \sum_{t=1}^T \Phi_{b,\ell} \left(\frac{L_n+1}{L_n}, \frac{M_n+1}{M_n}, \frac{t}{T} \right) L_n M_n t Q (1 + o_p(1)). \tag{S.12}
\end{aligned}$$

For I_3 , we have

$$\begin{aligned}
I_3 &= -\frac{1}{nT} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^T \left[\Phi_{b,\ell} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t+1}{T} \right) - \Phi_{b,\ell} \left(\frac{i_1}{L_n}, \frac{i_2+1}{M_n}, \frac{t+1}{T} \right) \right] \sum_{a_1=1}^{i_1} \sum_{a_2=1}^{i_2} \sum_{a_3=1}^t 1_{i_1,i_2} \tilde{Z}_{(i_1,i_2,t)} \tilde{X}'_{(i_1,i_2,t)} \\
&+ \frac{1}{nT} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \left[\Phi_{b,\ell} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{T+1}{T} \right) - \Phi_{b,\ell} \left(\frac{i_1}{L_n}, \frac{i_2+1}{M_n}, \frac{T+1}{T} \right) \right] \sum_{a_1=1}^{i_1} \sum_{a_2=1}^{i_2} \sum_{a_3=1}^T 1_{i_1,i_2} \tilde{Z}_{(i_1,i_2,t)} \tilde{X}'_{(i_1,i_2,t)} \\
&- \frac{1}{nT} \sum_{i_1=1}^{L_n} \sum_{t=1}^T \Phi_{b,\ell} \left(\frac{i_1}{L_n}, \frac{M_n+1}{M_n}, \frac{t+1}{T} \right) \sum_{a_1=1}^{i_1} \sum_{a_2=1}^{M_n} \sum_{a_3=1}^t 1_{i_1,i_2} \tilde{Z}_{(i_1,i_2,t)} \tilde{X}'_{(i_1,i_2,t)} \\
&+ \frac{1}{nT} \sum_{i_1=1}^{L_n} \Phi_{b,\ell} \left(\frac{i_1}{L_n}, \frac{M_n+1}{M_n}, \frac{T+1}{T} \right) \sum_{a_1=1}^{i_1} \sum_{a_2=1}^{M_n} \sum_{a_3=1}^T 1_{i_1,i_2} \tilde{Z}_{(i_1,i_2,t)} \tilde{X}'_{(i_1,i_2,t)}
\end{aligned}$$

$$\begin{aligned}
&= -\frac{1}{nT} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^T \left[\Phi_{b,\ell} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t+1}{T} \right) - \Phi_{b,\ell} \left(\frac{i_1}{L_n}, \frac{i_2+1}{M_n}, \frac{t+1}{T} \right) \right] i_1 i_2 t Q (1 + o_p(1)) \\
&+ \frac{1}{nT} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \left[\Phi_{b,\ell} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{T+1}{T} \right) - \Phi_{b,\ell} \left(\frac{i_1}{L_n}, \frac{i_2+1}{M_n}, \frac{T+1}{T} \right) \right] i_1 i_2 T Q (1 + o_p(1)) \\
&- \frac{1}{nT} \sum_{i_1=1}^{L_n} \sum_{t=1}^T \Phi_{b,\ell} \left(\frac{i_1}{L_n}, \frac{M_n+1}{M_n}, \frac{t+1}{T} \right) i_1 M_n t Q (1 + o_p(1)) \\
&+ \frac{1}{nT} \sum_{i_1=1}^{L_n} \Phi_{b,\ell} \left(\frac{i_1}{L_n}, \frac{M_n+1}{M_n}, \frac{T+1}{T} \right) i_1 M_n T Q (1 + o_p(1)). \tag{S.13}
\end{aligned}$$

For I_4 , we have

$$\begin{aligned}
I_4 &= \frac{1}{nT} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^T \left[\Phi_{b,\ell} \left(\frac{i_1+1}{L_n}, \frac{i_2}{M_n}, \frac{t+1}{T} \right) - \Phi_{b,\ell} \left(\frac{i_1+1}{L_n}, \frac{i_2+1}{M_n}, \frac{t+1}{T} \right) \right] i_1 i_2 t Q (1 + o_p(1)) \\
&- \frac{1}{nT} \sum_{i_2=1}^{M_n} \sum_{t=1}^T \left[\Phi_{b,\ell} \left(\frac{L_n+1}{L_n}, \frac{i_2}{M_n}, \frac{t+1}{T} \right) - \Phi_{b,\ell} \left(\frac{L_n+1}{L_n}, \frac{i_2+1}{M_n}, \frac{t+1}{T} \right) \right] L_n i_2 t Q (1 + o_p(1)) \\
&- \frac{1}{nT} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \left[\Phi_{b,\ell} \left(\frac{i_1+1}{L_n}, \frac{i_2}{M_n}, \frac{T+1}{T} \right) - \Phi_{b,\ell} \left(\frac{i_1+1}{L_n}, \frac{i_2+1}{M_n}, \frac{T+1}{T} \right) \right] i_1 i_2 T Q (1 + o_p(1)) \\
&+ \frac{1}{nT} \sum_{i_2=1}^{M_n} \left[\Phi_{b,\ell} \left(\frac{L_n+1}{L_n}, \frac{i_2}{M_n}, \frac{T+1}{T} \right) - \Phi_{b,\ell} \left(\frac{L_n+1}{L_n}, \frac{i_2+1}{M_n}, \frac{T+1}{T} \right) \right] L_n i_2 T Q (1 + o_p(1)) \\
&+ \frac{1}{nT} \sum_{i_1=1}^{L_n} \sum_{t=1}^T \Phi_{b,\ell} \left(\frac{i_1+1}{L_n}, \frac{M_n+1}{M_n}, \frac{t+1}{T} \right) i_1 M_n t Q (1 + o_p(1)) \\
&- \frac{1}{nT} \sum_{t=1}^T \Phi_{b,\ell} \left(\frac{L_n+1}{L_n}, \frac{M_n+1}{M_n}, \frac{t+1}{T} \right) L_n M_n t Q (1 + o_p(1)) \\
&- \frac{1}{nT} \sum_{i_1=1}^{L_n} \Phi_{b,\ell} \left(\frac{i_1+1}{L_n}, \frac{M_n+1}{M_n}, \frac{T+1}{T} \right) i_1 M_n T Q (1 + o_p(1)) \\
&+ \frac{1}{nT} \Phi_{b,\ell} \left(\frac{L_n+1}{L_n}, \frac{M_n+1}{M_n}, \frac{T+1}{T} \right) L_n M_n T Q (1 + o_p(1)). \tag{S.14}
\end{aligned}$$

From (S.11),(S.12),(S.13) and (S.14), we have

$$\begin{aligned}
& \frac{1}{nT} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^T \Phi_{b,\iota} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) 1_{i_1, i_2} \tilde{Z}_{(i_1, i_2, t)} \tilde{X}'_{(i_1, i_2, t)} \\
&= \frac{1}{nT} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^T \left[\Phi_{b,\iota} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) - \Phi_{b,\iota} \left(\frac{i_1+1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) \right. \\
&\quad - \Phi_{b,\iota} \left(\frac{i_1}{L_n}, \frac{i_2+1}{M_n}, \frac{t}{T} \right) - \Phi_{b,\iota} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t+1}{T} \right) + \Phi_{b,\iota} \left(\frac{i_1+1}{L_n}, \frac{i_2+1}{M_n}, \frac{t}{T} \right) + \Phi_{b,\iota} \left(\frac{i_1}{L_n}, \frac{i_2+1}{M_n}, \frac{t+1}{T} \right) \\
&\quad \left. + \Phi_{b,\iota} \left(\frac{i_1+1}{L_n}, \frac{i_2}{M_n}, \frac{t+1}{T} \right) - \Phi_{b,\iota} \left(\frac{i_1+1}{L_n}, \frac{i_2+1}{M_n}, \frac{t+1}{T} \right) \right] i_1 i_2 t Q (1 + o_p(1)) \\
&\quad + \frac{1}{nT} \sum_{i_2=1}^{M_n} \sum_{t=1}^T \left[\Phi_{b,\iota} \left(\frac{L_n+1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) - \Phi_{b,\iota} \left(\frac{L_n+1}{L_n}, \frac{i_2+1}{M_n}, \frac{t}{T} \right) \right. \\
&\quad \left. - \Phi_{b,\iota} \left(\frac{L_n+1}{L_n}, \frac{i_2}{M_n}, \frac{t+1}{T} \right) + \Phi_{b,\iota} \left(\frac{L_n+1}{L_n}, \frac{i_2+1}{M_n}, \frac{t+1}{T} \right) \right] L_n i_2 t Q (1 + o_p(1)) \\
&\quad + \frac{1}{nT} \sum_{i_1=1}^{L_n} \sum_{t=1}^T \left[\Phi_{b,\iota} \left(\frac{i_1}{L_n}, \frac{M_n+1}{M_n}, \frac{t}{T} \right) - \Phi_{b,\iota} \left(\frac{i_1+1}{L_n}, \frac{M_n+1}{M_n}, \frac{t}{T} \right) \right. \\
&\quad \left. - \Phi_{b,\iota} \left(\frac{i_1}{L_n}, \frac{M_n+1}{M_n}, \frac{t+1}{T} \right) + \Phi_{b,\iota} \left(\frac{i_1+1}{L_n}, \frac{M_n+1}{M_n}, \frac{t+1}{T} \right) \right] i_1 M_n t Q (1 + o_p(1)) \\
&\quad + \frac{1}{nT} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \left[\Phi_{b,\iota} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{T+1}{T} \right) - \Phi_{b,\iota} \left(\frac{i_1}{L_n}, \frac{i_2+1}{M_n}, \frac{T+1}{T} \right) \right. \\
&\quad \left. - \Phi_{b,\iota} \left(\frac{i_1+1}{L_n}, \frac{i_2}{M_n}, \frac{T+1}{T} \right) + \Phi_{b,\iota} \left(\frac{i_1+1}{L_n}, \frac{i_2+1}{M_n}, \frac{T+1}{T} \right) \right] i_1 i_2 T Q (1 + o_p(1)) \\
&\quad + \frac{1}{nT} \sum_{i_1=1}^{L_n} \left[\Phi_{b,\iota} \left(\frac{i_1}{L_n}, \frac{M_n+1}{M_n}, \frac{T+1}{T} \right) - \Phi_{b,\iota} \left(\frac{i_1+1}{L_n}, \frac{M_n+1}{M_n}, \frac{T+1}{T} \right) \right] i_1 M_n T Q (1 + o_p(1)) \\
&\quad + \frac{1}{nT} \sum_{i_2=1}^{M_n} \left[\Phi_{b,\iota} \left(\frac{L_n+1}{L_n}, \frac{i_2}{M_n}, \frac{T+1}{T} \right) - \Phi_{b,\iota} \left(\frac{L_n+1}{L_n}, \frac{i_2+1}{M_n}, \frac{T+1}{T} \right) \right] L_n i_2 T Q (1 + o_p(1)) \\
&\quad + \frac{1}{nT} \sum_{t=1}^T \left[\Phi_{b,\iota} \left(\frac{L_n+1}{L_n}, \frac{M_n+1}{M_n}, \frac{t}{T} \right) - \Phi_{b,\iota} \left(\frac{L_n+1}{L_n}, \frac{M_n+1}{M_n}, \frac{t+1}{T} \right) \right] L_n M_n t Q (1 + o_p(1)) \\
&\quad + \frac{1}{nT} \Phi_{b,\iota} \left(\frac{L_n+1}{L_n}, \frac{M_n+1}{M_n}, \frac{T+1}{T} \right) L_n M_n T Q (1 + o_p(1)) \\
&= A_1 + A_2 + A_3 + A_4 + A_5 + A_6 + A_7 + A_8.
\end{aligned}$$

After some algebra, we obtain

$$\begin{aligned}
A_1 &= \left(\frac{1}{nT} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^T \Phi_{b,\iota} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) Q - \frac{1}{nT} \sum_{i_2=1}^{M_n} \sum_{t=1}^T \Phi_{b,\iota} \left(\frac{L_n+1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) L_n Q \right. \\
&\quad - \frac{1}{nT} \sum_{i_1=1}^{L_n} \sum_{t=1}^T \Phi_{b,\iota} \left(\frac{i_1}{L_n}, \frac{M_n+1}{M_n}, \frac{t}{T} \right) M_n Q - \frac{1}{nT} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \Phi_{b,\iota} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{T+1}{T} \right) TQ \\
&\quad + \frac{1}{nT} \sum_{i_1=1}^{L_n} \Phi_{b,\iota} \left(\frac{i_1}{L_n}, \frac{M_n+1}{M_n}, \frac{T+1}{T} \right) M_n TQ + \frac{1}{nT} \sum_{i_2=1}^{M_n} \Phi_{b,\iota} \left(\frac{L_n+1}{L_n}, \frac{i_2}{M_n}, \frac{T+1}{T} \right) L_n TQ \\
&\quad \left. + \frac{1}{nT} \sum_{t=1}^T \Phi_{b,\iota} \left(\frac{L_n+1}{L_n}, \frac{M_n+1}{M_n}, \frac{t}{T} \right) L_n M_n Q - \frac{1}{nT} \Phi_{b,\iota} \left(\frac{L_n+1}{L_n}, \frac{M_n+1}{M_n}, \frac{T+1}{T} \right) L_n M_n TQ \right) (1 + o_p(1)), \\
A_2 &= \left(\frac{1}{nT} \sum_{i_2=1}^{M_n} \sum_{t=1}^T \Phi_{b,\iota} \left(\frac{L_n+1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) L_n Q - \frac{1}{nT} \sum_{i_2=1}^{M_n} \Phi_{b,\iota} \left(\frac{L_n+1}{L_n}, \frac{i_2}{M_n}, \frac{T+1}{T} \right) L_n TQ \right. \\
&\quad \left. - \frac{1}{nT} \sum_{t=1}^T \Phi_{b,\iota} \left(\frac{L_n+1}{L_n}, \frac{M_n+1}{M_n}, \frac{t}{T} \right) L_n M_n Q + \frac{1}{nT} \Phi_{b,\iota} \left(\frac{L_n+1}{L_n}, \frac{M_n+1}{M_n}, \frac{T+1}{T} \right) L_n M_n TQ \right) (1 + o_p(1)), \\
A_3 &= \left(\frac{1}{nT} \sum_{i_1=1}^{L_n} \sum_{t=1}^T \Phi_{b,\iota} \left(\frac{i_1}{L_n}, \frac{M_n+1}{M_n}, \frac{t}{T} \right) M_n Q - \frac{1}{nT} \sum_{i_1=1}^{L_n} \Phi_{b,\iota} \left(\frac{i_1}{L_n}, \frac{M_n+1}{M_n}, \frac{T+1}{T} \right) M_n TQ \right. \\
&\quad \left. - \frac{1}{nT} \sum_{t=1}^T \Phi_{b,\iota} \left(\frac{L_n+1}{L_n}, \frac{M_n+1}{M_n}, \frac{t}{T} \right) L_n M_n Q + \frac{1}{nT} \Phi_{b,\iota} \left(\frac{L_n+1}{L_n}, \frac{M_n+1}{M_n}, \frac{T+1}{T} \right) L_n M_n TQ \right) (1 + o_p(1)), \\
A_4 &= \left(\frac{1}{nT} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \Phi_{b,\iota} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{T+1}{T} \right) TQ - \frac{1}{nT} \sum_{i_2=1}^{M_n} \Phi_{b,\iota} \left(\frac{L_n+1}{L_n}, \frac{i_2}{M_n}, \frac{T+1}{T} \right) L_n TQ \right. \\
&\quad \left. - \frac{1}{nT} \sum_{i_1=1}^{L_n} \Phi_{b,\iota} \left(\frac{i_1}{L_n}, \frac{M_n+1}{M_n}, \frac{T+1}{T} \right) M_n TQ + \frac{1}{nT} \Phi_{b,\iota} \left(\frac{L_n+1}{L_n}, \frac{M_n+1}{M_n}, \frac{T+1}{T} \right) L_n M_n TQ \right) (1 + o_p(1)), \\
A_5 &= \left(\frac{1}{nT} \sum_{i_1=1}^{L_n} \Phi_{b,\iota} \left(\frac{i_1}{L_n}, \frac{M_n+1}{M_n}, \frac{T+1}{T} \right) M_n TQ - \frac{1}{nT} \Phi_{b,\iota} \left(\frac{L_n+1}{L_n}, \frac{M_n+1}{M_n}, \frac{T+1}{T} \right) L_n M_n TQ \right) (1 + o_p(1)), \\
A_6 &= \left(\frac{1}{nT} \sum_{i_2=1}^{M_n} \Phi_{b,\iota} \left(\frac{L_n+1}{L_n}, \frac{i_2}{M_n}, \frac{T+1}{T} \right) L_n TQ - \frac{1}{nT} \Phi_{b,\iota} \left(\frac{L_n+1}{L_n}, \frac{M_n+1}{M_n}, \frac{T+1}{T} \right) L_n M_n TQ \right) (1 + o_p(1)), \\
A_7 &= \left(\frac{1}{nT} \sum_{t=1}^T \Phi_{b,\iota} \left(\frac{L_n+1}{L_n}, \frac{M_n+1}{M_n}, \frac{t}{T} \right) L_n M_n Q - \frac{1}{nT} \Phi_{b,\iota} \left(\frac{L_n+1}{L_n}, \frac{M_n+1}{M_n}, \frac{T+1}{T} \right) L_n M_n TQ \right) (1 + o_p(1)).
\end{aligned}$$

Therefore,

$$\begin{aligned}
& \frac{1}{nT} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^T \Phi_{b,\iota} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) 1_{i_1, i_2} \tilde{Z}_{(i_1, i_2, t)} \tilde{X}'_{(i_1, i_2, t)} \\
&= \frac{1}{nT} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^T \Phi_{b,\iota} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) Q (1 + o_p(1)) \\
&= \frac{1}{nT} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^T \Phi_{b,\iota} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) Q + o_p(1), \tag{S.15}
\end{aligned}$$

where the second equality holds because

$$\frac{1}{nT} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^T \left| \Phi_{b,\iota} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) \right| Q = O(1).$$

Applying (S.15) to (S.10), we finally have

$$\begin{aligned}
\mathfrak{X}_{\iota, nT} &= \frac{1}{\sqrt{nT}} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^T \Phi_{b,\iota} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) V_{(i_1, i_2, t)}^* \\
&\quad - \frac{1}{\sqrt{nT}} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^T \Phi_{b,\iota} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) \left(\frac{1}{nT} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^T V_{(i_1, i_2, t)}^* \right) + o_p(1) \\
&= \frac{1}{\sqrt{nT}} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^T \Phi_{b,\iota} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) \left(V_{(i_1, i_2, t)}^* - \frac{1}{nT} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^T V_{(i_1, i_2, t)}^* \right) + o_p(1) \\
&\stackrel{a}{\approx} \Lambda \frac{1}{\sqrt{nT}} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^T \Phi_{b,\iota} \left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T} \right) \left(e_{(i_1, i_2, t)}^* - \frac{1}{nT} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^T e_{(i_1, i_2, t)}^* \right)
\end{aligned}$$

by Assumption F3. The proof is complete. ■

Proof of Lemma 3

Let $i = \sqrt{-1}$. To simplify the notation, we use \int to stand for $\int \int \dots \int$ when there is no confusion. Under Assumption F1, let's denote

$$\begin{aligned}
\mathbb{K}_1((x_1, x_2), (y_1, y_2)) &= K_1(d_{(x_1, x_2), (y_1, y_2)}), \\
\mathbb{K}_{1b}((x_1, x_2), (y_1, y_2)) &= \mathbb{K}_1((x_1/b_1, x_2/b_2), (y_1/b_1, y_2/b_2)), \\
\mathbb{K}_2(x_3, y_3) &= K_2(d_{x_3, y_3}), \quad \mathbb{K}_{2b}(x_3, y_3) = \mathbb{K}_2(x_3/b_3, y_3/b_3).
\end{aligned}$$

We also define

$$c_{1S} = \int_{-1}^1 \int_{-1}^1 \mathbb{K}_1((x_1, x_2), (x_1 - z_1, x_2 - z_2)) dz_1 dz_2, \quad c_{1T} = \int_{-1}^1 \mathbb{K}_2(x_3, x_3 - z_3) dz_3,$$

and

$$\begin{aligned}
\mathcal{K}_1(\lambda_1, \lambda_2) &= \frac{1}{4\pi^2} \int_{-1}^1 \int_{-1}^1 \mathbb{K}_1((x_1, x_2), (x_1 - z_1, x_2 - z_2)) \exp(-i\lambda_1 z_1 - i\lambda_2 z_2) dz_1 dz_2, \\
\mathcal{K}_1^{(2)}(\lambda_1, \lambda_2) &= \frac{1}{4\pi^2} \int_{-1}^1 \int_{-1}^1 \mathbb{K}_1^2((x_1, x_2), (x_1 - z_1, x_2 - z_2)) \exp(-i\lambda_1 z_1 - i\lambda_2 z_2) dz_1 dz_2, \\
\mathcal{K}_2(\lambda_3) &= \frac{1}{2\pi} \int_{-1}^1 \mathbb{K}_2(x_3, x_3 - z_3) \exp(-i\lambda_3 z_3) dz_3, \\
\mathcal{K}_2^{(2)}(\lambda_3) &= \frac{1}{2\pi} \int_{-1}^1 \mathbb{K}_2^2(x_3, x_3 - z_3) \exp(-i\lambda_3 z_3) dz_3.
\end{aligned}$$

Then, we have

$$\begin{aligned}
&\mathbb{K}_b((x_1, x_2, x_3), (y_1, y_2, y_3)) \\
&= \mathbb{K}_{1b}((x_1, x_2), (y_1, y_2)) \mathbb{K}_{2b}(x_3, y_3) \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{K}_1(\lambda_1, \lambda_2) \mathcal{K}_2(\lambda_3) \exp\left(i\lambda_1 \frac{x_1 - y_1}{b_1} + i\lambda_2 \frac{x_2 - y_2}{b_2} + i\lambda_3 \frac{x_3 - y_3}{b_3}\right) d\lambda_1 d\lambda_2 d\lambda_3, \\
&\mathbb{K}_b^2((x_1, x_2, x_3), (y_1, y_2, y_3)) \\
&= \mathbb{K}_{1b}^2((x_1, x_2), (y_1, y_2)) \mathbb{K}_{2b}^2(x_3, y_3) \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{K}_1^{(2)}(\lambda_1, \lambda_2) \mathcal{K}_2^{(2)}(\lambda_3) \exp\left(i\lambda_1 \frac{x_1 - y_1}{b_1} + i\lambda_2 \frac{x_2 - y_2}{b_2} + i\lambda_3 \frac{x_3 - y_3}{b_3}\right) d\lambda_1 d\lambda_2 d\lambda_3.
\end{aligned}$$

Note that

$$\begin{aligned}
\mu_1 &= \int_0^1 \int_0^1 \int_0^1 \mathbb{K}_b^*((x_1, x_2, x_3), (x_1, x_2, x_3)) dx_1 dx_2 dx_3 \\
&= 1 - \int_0^1 \mathbb{K}_b((x_1, x_2, x_3), (y_1, y_2, y_3)) dx_1 dy_1 dx_2 dy_2 dx_3 dy_3.
\end{aligned} \tag{S.16}$$

Now

$$\begin{aligned}
&\int_0^1 \mathbb{K}_b((x_1, x_2, x_3), (y_1, y_2, y_3)) dx_1 dy_1 dx_2 dy_2 dx_3 dy_3 \\
&= \int_{-\infty}^{\infty} \mathcal{K}_1(\lambda_1, \lambda_2) \mathcal{K}_2(\lambda_3) \int_0^1 \exp\left(i\lambda_1 \frac{x_1 - y_1}{b_1} + i\lambda_2 \frac{x_2 - y_2}{b_2} + i\lambda_3 \frac{x_3 - y_3}{b_3}\right) dx_1 dy_1 dx_2 dy_2 dx_3 dy_3 d\lambda_1 d\lambda_2 d\lambda_3 \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{K}_1(\lambda_1, \lambda_2) \left(\frac{b_1 b_2}{\lambda_1 \lambda_2}\right)^2 \left[\left(1 - \cos\left(\frac{\lambda_1}{b_1}\right)\right)^2 + \sin^2\left(\frac{\lambda_1}{b_1}\right)\right] \\
&\times \left[\left(1 - \cos\left(\frac{\lambda_2}{b_2}\right)\right)^2 + \sin^2\left(\frac{\lambda_2}{b_2}\right)\right] d\lambda_1 d\lambda_2 \int_{-\infty}^{\infty} \mathcal{K}_2(\lambda_3) \left(\frac{b_3}{\lambda_3}\right)^2 \left[\left(1 - \cos\left(\frac{\lambda_3}{b_3}\right)\right)^2 + \sin^2\left(\frac{\lambda_3}{b_3}\right)\right] d\lambda_3 \\
&= \left[(b_1 b_2)^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{K}_1(\lambda_1, \lambda_2) \left(\frac{\sin\left(\frac{\lambda_1}{2b_1}\right)}{\frac{\lambda_1}{2}}\right)^2 \left(\frac{\sin\left(\frac{\lambda_2}{2b_2}\right)}{\frac{\lambda_2}{2}}\right)^2 d\lambda_1 d\lambda_2\right] (b_3 c_{1T} + o(b_3)), \tag{S.17}
\end{aligned}$$

where the last equality holds by Sun (2010).

$$\begin{aligned}
& (b_1 b_2)^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{K}_1(\lambda_1, \lambda_2) \left(\frac{\sin\left(\frac{\lambda_1}{2b_1}\right)}{\frac{\lambda_1}{2}} \right)^2 \left(\frac{\sin\left(\frac{\lambda_2}{2b_2}\right)}{\frac{\lambda_2}{2}} \right)^2 d\lambda_1 d\lambda_2 \\
&= (b_1 b_2)^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [\mathcal{K}_1(\lambda_1, \lambda_2) - \mathcal{K}_1(\lambda_1, 0)] \left(\frac{\sin\left(\frac{\lambda_1}{2b_1}\right)}{\frac{\lambda_1}{2}} \right)^2 \left(\frac{\sin\left(\frac{\lambda_2}{2b_2}\right)}{\frac{\lambda_2}{2}} \right)^2 d\lambda_1 d\lambda_2 \\
&+ (b_1 b_2)^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [\mathcal{K}_1(\lambda_1, 0) - \mathcal{K}_1(0, 0)] \left(\frac{\sin\left(\frac{\lambda_1}{2b_1}\right)}{\frac{\lambda_1}{2}} \right)^2 \left(\frac{\sin\left(\frac{\lambda_2}{2b_2}\right)}{\frac{\lambda_2}{2}} \right)^2 d\lambda_1 d\lambda_2 \\
&+ (b_1 b_2)^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{K}_1(0, 0) \left(\frac{\sin\left(\frac{\lambda_1}{2b_1}\right)}{\frac{\lambda_1}{2}} \right)^2 \left(\frac{\sin\left(\frac{\lambda_2}{2b_2}\right)}{\frac{\lambda_2}{2}} \right)^2 d\lambda_1 d\lambda_2. \tag{S.18}
\end{aligned}$$

The first term in (S.18) is

$$\begin{aligned}
& 4(b_1 b_2)^2 \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \left[\frac{\mathcal{K}_1(\lambda_1, \lambda_2) - \mathcal{K}_1(\lambda_1, 0)}{\lambda_2^2} \right] \left(\sin^2\left(\frac{\lambda_2}{2b_2}\right) - \frac{1}{2} \right) d\lambda_2 \right] \left(\frac{\sin\left(\frac{\lambda_1}{2b_1}\right)}{\frac{\lambda_1}{2}} \right)^2 d\lambda_1 \\
&+ 2(b_1 b_2)^2 \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \left(\frac{\mathcal{K}_1(\lambda_1, \lambda_2) - \mathcal{K}_1(\lambda_1, 0)}{\lambda_2^2} \right) d\lambda_2 \right] \left(\frac{\sin\left(\frac{\lambda_1}{2b_1}\right)}{\frac{\lambda_1}{2}} \right)^2 d\lambda_1 \\
&= -2(b_1 b_2)^2 \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \left(\frac{\mathcal{K}_1(\lambda_1, \lambda_2) - \mathcal{K}_1(\lambda_1, 0)}{\lambda_2^2} \right) \left(\frac{\sin\left(\frac{\lambda_1}{2b_1}\right)}{\frac{\lambda_1}{2}} \right)^2 d\lambda_1 \right] \cos\left(\frac{\lambda_2}{b_2}\right) d\lambda_2 \\
&+ 2(b_1 b_2)^2 \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \left(\frac{\mathcal{K}_1(\lambda_1, \lambda_2) - \mathcal{K}_1(\lambda_1, 0)}{\lambda_2^2} \right) d\lambda_2 \right] \left(\frac{\sin\left(\frac{\lambda_1}{2b_1}\right)}{\frac{\lambda_1}{2}} \right)^2 d\lambda_1 \\
&= O(b_1 b_2^2), \tag{S.19}
\end{aligned}$$

because

$$\begin{aligned}
& 2(b_1 b_2)^2 \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \left(\frac{\mathcal{K}_1(\lambda_1, \lambda_2) - \mathcal{K}_1(\lambda_1, 0)}{\lambda_2^2} \right) d\lambda_2 \right] \left(\frac{\sin\left(\frac{\lambda_1}{2b_1}\right)}{\frac{\lambda_1}{2}} \right)^2 d\lambda_1 \\
&= \frac{(b_1 b_2)^2}{2\pi^2} \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \int_{-1}^1 \int_{-1}^1 \left(\mathbb{K}_1((x_1, x_2), (x_1 - z_1, x_2 - z_2)) \exp(-i\lambda_1 z_1) \frac{\exp(-i\lambda_2 z_2) - 1}{\lambda_2^2} \right) dz_1 dz_2 d\lambda_2 \right] \\
&\quad \times \left(\frac{\sin\left(\frac{\lambda_1}{2b_1}\right)}{\frac{\lambda_1}{2}} \right)^2 d\lambda_1 \\
&= \frac{(b_1 b_2)^2}{2\pi^2} \int_{-\infty}^{\infty} \left[\int_{-1}^1 \int_{-1}^1 \left(\mathbb{K}_1((x_1, x_2), (x_1 - z_1, x_2 - z_2)) \exp(-i\lambda_1 z_1) \int_{-\infty}^{\infty} \frac{\cos(\lambda_2 z_2) - 1}{\lambda_2^2} d\lambda_2 \right) dz_1 dz_2 \right] \\
&\quad \times \left(\frac{\sin\left(\frac{\lambda_1}{2b_1}\right)}{\frac{\lambda_1}{2}} \right)^2 d\lambda_1 \\
&= -\frac{(b_1 b_2)^2}{\pi^2} \int_{-\infty}^{\infty} \left[\int_{-1}^1 \int_{-1}^1 \left(\mathbb{K}_1((x_1, x_2), (x_1 - z_1, x_2 - z_2)) \exp(-i\lambda_1 z_1) \int_{-\infty}^{\infty} \frac{\sin^2(\lambda_2 z_2/2)}{\lambda_2^2} d\lambda_2 \right) dz_1 dz_2 \right] \\
&\quad \times \left(\frac{\sin\left(\frac{\lambda_1}{2b_1}\right)}{\frac{\lambda_1}{2}} \right)^2 d\lambda_1 \\
&= -\frac{1}{2} \frac{(b_1 b_2)^2}{\pi} \int_{-\infty}^{\infty} \left[\int_{-1}^1 \int_{-1}^1 \mathbb{K}_1((x_1, x_2), (x_1 - z_1, x_2 - z_2)) |z_2| (\exp(-i\lambda_1 z_1) - 1) dz_1 dz_2 \right] \left(\frac{\sin\left(\frac{\lambda_1}{2b_1}\right)}{\frac{\lambda_1}{2}} \right)^2 d\lambda_1 \\
&\quad - \frac{1}{2} \frac{(b_1 b_2)^2}{\pi} \int_{-\infty}^{\infty} \left[\int_{-1}^1 \int_{-1}^1 \mathbb{K}_1((x_1, x_2), (x_1 - z_1, x_2 - z_2)) |z_2| dz_1 dz_2 \right] \left(\frac{\sin\left(\frac{\lambda_1}{2b_1}\right)}{\frac{\lambda_1}{2}} \right)^2 d\lambda_1 \\
&= -2 \frac{(b_1 b_2)^2}{\pi} \int_{-\infty}^{\infty} \left[\int_{-1}^1 \int_{-1}^1 \mathbb{K}_1((x_1, x_2), (x_1 - z_1, x_2 - z_2)) |z_2| \left(\frac{\exp(-i\lambda_1 z_1) - 1}{\lambda_1^2} \right) dz_1 dz_2 \right] \sin^2\left(\frac{\lambda_1}{2b_1}\right) d\lambda_1 \\
&\quad - \frac{1}{2} \frac{b_2^2}{\pi} \left[\int_{-1}^1 \int_{-1}^1 \mathbb{K}_1((x_1, x_2), (x_1 - z_1, x_2 - z_2)) |z_2| dz_1 dz_2 \right] \int_{-\infty}^{\infty} \left(\frac{\sin\left(\frac{\lambda_1}{2b_1}\right)}{\frac{\lambda_1}{2b_1}} \right)^2 d\lambda_1 \\
&= -2 \frac{(b_1 b_2)^2}{\pi} \int_{-\infty}^{\infty} \left[\int_{-1}^1 \int_{-1}^1 \mathbb{K}_1((x_1, x_2), (x_1 - z_1, x_2 - z_2)) |z_2| \frac{\exp(-i\lambda_1 z_1) - 1}{\lambda_1^2} dz_1 dz_2 \right] \left(\sin^2\left(\frac{\lambda_1}{2b_1}\right) - \frac{1}{2} \right) d\lambda_1 \\
&\quad - \frac{(b_1 b_2)^2}{\pi} \int_{-1}^1 \int_{-1}^1 \mathbb{K}_1((x_1, x_2), (x_1 - z_1, x_2 - z_2)) |z_2| dz_2 \int_{-\infty}^{\infty} \left(\frac{\exp(-i\lambda_1 z_1) - 1}{\lambda_1^2} \right) d\lambda_1 dz_1 \\
&\quad - \frac{1}{2} \frac{b_2^2}{\pi} \left[\int_{-1}^1 \int_{-1}^1 \mathbb{K}_1((x_1, x_2), (x_1 - z_1, x_2 - z_2)) |z_2| dz_1 dz_2 \right] \int_{-\infty}^{\infty} \left(\frac{\sin\left(\frac{\lambda_1}{2b_1}\right)}{\frac{\lambda_1}{2b_1}} \right)^2 d\lambda_1
\end{aligned}$$

$$\begin{aligned}
&= \frac{(b_1 b_2)^2}{\pi} \int_{-\infty}^{\infty} \left[\int_{-1}^1 \int_{-1}^1 \mathbb{K}_1((x_1, x_2), (x_1 - z_1, x_2 - z_2)) |z_2| \left(\frac{\exp(-i\lambda_1 z_1) - 1}{\lambda_1^2} \right) dz_1 dz_2 \right] \cos\left(\frac{\lambda_1}{b_1}\right) d\lambda_1 \\
&+ 2 \frac{(b_1 b_2)^2}{\pi} \int_{-1}^1 \int_{-1}^1 \mathbb{K}_1((x_1, x_2), (x_1 - z_1, x_2 - z_2)) |z_2| dz_2 \int_{-\infty}^{\infty} \left(\frac{\sin^2(\lambda_1 z_1/2)}{\lambda_1^2} \right) d\lambda_1 dz_1 \\
&- b_1 b_2^2 \int_{-1}^1 \int_{-1}^1 \mathbb{K}_1((x_1, x_2), (x_1 - z_1, x_2 - z_2)) |z_2| dz_1 dz_2 \\
&= (b_1 b_2)^2 \int_{-1}^1 \int_{-1}^1 \mathbb{K}_1((x_1, x_2), (x_1 - z_1, x_2 - z_2)) |z_1| |z_2| dz_1 dz_2 \\
&- b_1 b_2^2 \int_{-1}^1 \int_{-1}^1 \mathbb{K}_1((x_1, x_2), (x_1 - z_1, x_2 - z_2)) |z_2| dz_1 dz_2 + o\left((b_1 b_2)^2\right)
\end{aligned}$$

as $b_1, b_2 \rightarrow 0$, where we have used the Riemann-Lebesgue lemma, and this implies

$$-2 (b_1 b_2)^2 \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \left(\frac{\mathcal{K}_1(\lambda_1, \lambda_2) - \mathcal{K}_1(\lambda_1, 0)}{\lambda_2^2} \right) \left(\frac{\sin\left(\frac{\lambda_1}{2b_1}\right)}{\frac{\lambda_1}{2}} \right)^2 d\lambda_1 \right] \cos\left(\frac{\lambda_2}{b_2}\right) d\lambda_2 = o(b_1 b_2^2)$$

by the Riemann-Lebesgue lemma.

The second term in (S.18) is

$$\begin{aligned}
&b_1 b_2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [\mathcal{K}_1(\lambda_1, 0) - \mathcal{K}_1(0, 0)] b_1 b_2 \left(\frac{\sin\left(\frac{\lambda_1}{2b_1}\right)}{\frac{\lambda_1}{2}} \right)^2 \left(\frac{\sin\left(\frac{\lambda_2}{2b_2}\right)}{\frac{\lambda_2}{2}} \right)^2 d\lambda_1 d\lambda_2 \\
&= 4b_1^2 \int_{-\infty}^{\infty} \left[\frac{\mathcal{K}_1(\lambda_1, 0) - \mathcal{K}_1(0, 0)}{\lambda_1^2} \right] \sin^2\left(\frac{\lambda_1}{2b_1}\right) d\lambda_1 \int_{-\infty}^{\infty} \left(\frac{\sin\left(\frac{\lambda_2}{2b_2}\right)}{\frac{\lambda_2}{2}} \right)^2 d\lambda_2 \\
&= 8\pi b_1^2 b_2 \int_{-\infty}^{\infty} \left[\frac{\mathcal{K}_1(\lambda_1, 0) - \mathcal{K}_1(0, 0)}{\lambda_1^2} \right] \left(\sin^2\left(\frac{\lambda_1}{2b_1}\right) - \frac{1}{2} \right) d\lambda_1 \\
&+ 4\pi b_1^2 b_2 \int_{-\infty}^{\infty} \frac{\mathcal{K}_1(\lambda_1, 0) - \mathcal{K}_1(0, 0)}{\lambda_1^2} d\lambda_1 \tag{S.20}
\end{aligned}$$

$$\begin{aligned}
&= -4\pi b_1^2 b_2 \int_{-\infty}^{\infty} \left[\frac{\mathcal{K}_1(\lambda_1, 0) - \mathcal{K}_1(0, 0)}{\lambda_1^2} \right] \cos\left(\frac{\lambda_1}{b_1}\right) d\lambda_1 \\
&+ \frac{b_1^2 b_2}{\pi} \int_{-1}^1 \int_{-1}^1 \mathbb{K}_1((x_1, x_2), (x_1 - z_1, x_2 - z_2)) \int_{-\infty}^{\infty} \frac{\exp(-i\lambda_1 z_1) - 1}{\lambda_1^2} d\lambda_1 dz_1 dz_2 \\
&= \frac{b_1^2 b_2}{\pi} \int_{-1}^1 \int_{-1}^1 \mathbb{K}_1((x_1, x_2), (x_1 - z_1, x_2 - z_2)) \int_{-\infty}^{\infty} \frac{\cos(\lambda_1 z_1) - 1}{\lambda_1^2} d\lambda_1 dz_1 dz_2 + o(b_1^2 b_2) \\
&= -2 \frac{b_1^2 b_2}{\pi} \int_{-1}^1 \int_{-1}^1 \mathbb{K}_1((x_1, x_2), (x_1 - z_1, x_2 - z_2)) \int_{-\infty}^{\infty} \frac{\sin^2(\lambda_1 z_1/2)}{\lambda_1^2} d\lambda_1 dz_1 dz_2 + o(b_1^2 b_2) \\
&= -b_1^2 b_2 \int_{-1}^1 \int_{-1}^1 \mathbb{K}_1((x_1, x_2), (x_1 - z_1, x_2 - z_2)) |z_1| dz_1 dz_2 + o(b_1^2 b_2) = o(b_1 b_2) \tag{S.21}
\end{aligned}$$

The last term in (S.18) is

$$\begin{aligned} \mathcal{K}_1(0,0) & \int_{-\infty}^{\infty} \left(\frac{\sin\left(\frac{\lambda_1}{2b_1}\right)}{\frac{\lambda_1}{2b_1}} \right)^2 d\lambda_1 \int_{-\infty}^{\infty} \left(\frac{\sin\left(\frac{\lambda_2}{2b_2}\right)}{\frac{\lambda_2}{2b_2}} \right)^2 d\lambda_2 \\ & = b_1 b_2 \int_{-1}^1 \int_{-1}^1 \mathbb{K}_1((x_1, x_2), (x_1 - z_1, x_2 - z_2)) dz_1 dz_2. \end{aligned} \quad (\text{S.22})$$

With (S.19), (S.20) and (S.22), (S.18) equals

$$b_1 b_2 \int_{-1}^1 \int_{-1}^1 \mathbb{K}_1((x_1, x_2), (x_1 - z_1, x_2 - z_2)) dz_1 dz_2 + o(b_1 b_2) = b_1 b_2 c_{1S} + o(b_1 b_2)$$

Plugging this into (S.17), we have

$$\begin{aligned} & \int_0^1 \mathbb{K}_b((x_1, x_2, x_3), (y_1, y_2, y_3)) dx_1 dy_1 dx_2 dy_2 dx_3 dy_3 \\ & = b_1 b_2 b_3 c_1 + o(b_1 b_2 b_3), \end{aligned} \quad (\text{S.23})$$

and from (S.16) finally

$$\mu_1 = 1 - b_1 b_2 b_3 c_1 + o(b_1 b_2 b_3).$$

Next, note that

$$\begin{aligned} \mu_2 & = \sum_{k=1}^{\infty} (\lambda_k^*)^2 \\ & = \int_0^1 [\mathbb{K}_b^*((x_1, x_2, x_3), (y_1, y_2, y_3))]^2 dx_1 dy_1 dx_2 dy_2 dx_3 dy_3 \\ & = \left[\int_0^1 \mathbb{K}_b((x_1, x_2, x_3), (y_1, y_2, y_3)) dx_1 dy_1 dx_2 dy_2 dx_3 dy_3 \right]^2 \\ & + \int_0^1 \mathbb{K}_b^2((x_1, x_2, x_3), (y_1, y_2, y_3)) dx_1 dy_1 dx_2 dy_2 dx_3 dy_3 \\ & - 2 \int_0^1 \mathbb{K}_b((x_1, x_2, x_3), (r_1, r_2, r_3)) \mathbb{K}_b((x_1, x_2, x_3), (v_1, v_2, v_3)) dx_1 dx_2 dx_3 dr_1 dr_2 dr_3 dv_1 dv_2 dv_3. \end{aligned} \quad (\text{S.24})$$

It is straightforward from (S.23) that

$$\int_0^1 \mathbb{K}_b^2((x_1, x_2, x_3), (y_1, y_2, y_3)) dx_1 dx_2 dy_1 dy_2 dz_1 dz_2 = b_1 b_2 b_3 c_2 + o(b_1 b_2 b_3). \quad (\text{S.25})$$

For the last term of (S.25),

$$\begin{aligned} & \int_0^1 \mathbb{K}_b((x_1, x_2, x_3), (r_1, r_2, r_3)) \mathbb{K}_b((x_1, x_2, x_3), (v_1, v_2, v_3)) dx_1 dx_2 dx_3 dr_1 dr_2 dr_3 dv_1 dv_2 dv_3 \\ & = \int_0^1 \mathbb{K}_{1b}((x_1, x_2), (r_1, r_2)) \mathbb{K}_{1b}((x_1, x_2), (v_1, v_2)) dx_1 dx_2 dr_1 dr_2 dv_1 dv_2 \\ & \times \int_0^1 \mathbb{K}_{2b}(x_3, r_3) \mathbb{K}_{2b}(x_3, v_3) dx_3 dr_3 dv_3 \\ & = \left[\int_0^1 \mathbb{K}_{1b}((x_1, x_2), (r_1, r_2)) \mathbb{K}_{1b}((x_1, x_2), (v_1, v_2)) dx_1 dx_2 dr_1 dr_2 dv_1 dv_2 \right] O(b_3^2) \end{aligned} \quad (\text{S.26})$$

where the last equality holds by Sun (2010). As

$$\begin{aligned}
& \int_0^1 \int_0^1 \mathbb{K}_{1b}((x_1, x_2), (r_1, r_2)) dr_1 dr_2 \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{K}_1(\lambda_1, \lambda_2) \int_0^1 \int_0^1 \exp\left(i\lambda_1 \frac{(x_1 - r_1)}{b_1} + i\lambda_2 \frac{(x_2 - r_2)}{b_2}\right) dr_1 dr_2 d\lambda_1 d\lambda_2 \\
&= \frac{1}{2} \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \mathcal{K}_1(\lambda_1, \lambda_2) \int_0^1 \left(\exp\left(i\lambda_1 \frac{(x_1 - r_1)}{b_1}\right) + \exp\left(-i\lambda_1 \frac{(x_1 - r_1)}{b_1}\right) \right) dr_1 d\lambda_1 \right] \\
&\quad \times \int_0^1 \exp\left(i\lambda_2 \frac{(x_2 - r_2)}{b_2}\right) dr_2 d\lambda_2 \\
&= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \mathcal{K}_1(\lambda_1, \lambda_2) \int_0^1 \cos\left(\frac{\lambda_1 (x_1 - r_1)}{b_1}\right) dr_1 d\lambda_1 \right] \int_0^1 \exp\left(i\lambda_2 \frac{(x_2 - r_2)}{b_2}\right) dr_2 d\lambda_2 \\
&= -b_1 \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \mathcal{K}_1(\lambda_1, \lambda_2) \frac{1}{\lambda_1} \left(\sin\left(\frac{\lambda_1 (x_1 - 1)}{b_1}\right) - \sin\left(\frac{\lambda_1 x_1}{b_1}\right) \right) d\lambda_1 \right] \int_0^1 \exp\left(i\lambda_2 \frac{(x_2 - r_2)}{b_2}\right) dr_2 d\lambda_2 \\
&= -b_1 \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \mathcal{K}_1(\varphi b_1, \lambda_2) \frac{1}{\varphi} (\sin(\varphi(x_1 - 1)) - \sin(\varphi x_1)) d\varphi \right] \int_0^1 \cos\left(\frac{\lambda_2 (x_2 - r_2)}{b_2}\right) dr_2 d\lambda_2 \\
&= b_1 b_2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{K}_1(\varphi b_1, \psi b_2) \frac{1}{\varphi \psi} (\sin(\varphi(x_1 - 1)) - \sin(\varphi x_1)) (\sin(\psi(x_2 - 1)) - \sin(\psi x_2)) d\varphi d\psi,
\end{aligned}$$

where the second equality holds by the symmetry condition of $\mathbb{K}_1((x_1, x_2), (y_1, y_2))$ in Assumption F2(i). With this, we obtain

$$\begin{aligned}
& \int_0^1 \mathbb{K}_{1b}((x_1, x_2), (r_1, r_2)) \mathbb{K}_{1b}((x_1, x_2), (v_1, v_2)) dx_1 dx_2 dr_1 dr_2 dv_1 dv_2 \\
&= (b_1 b_2)^2 \int_0^1 \int_0^1 \left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{K}_1(\varphi b_1, \psi b_2) \frac{1}{\varphi \psi} (\sin(\varphi(x_1 - 1)) - \sin(\varphi x_1)) (\sin(\psi(x_2 - 1)) \right. \\
&\quad \left. - \sin(\psi x_2)) d\varphi d\psi \right]^2 dx_1 dx_2 \\
&= (b_1 b_2)^2 \mathcal{K}_1^2(0, 0) \left[\int_0^1 \left(\int_{-\infty}^{\infty} \frac{1}{\varphi} (\sin(\varphi(x_1 - 1)) - \sin(\varphi x_1)) d\varphi \right)^2 dx_1 \right]^2 (1 + o(1)) \\
&= (b_1 b_2)^2 \mathcal{K}_1^2(0, 0) \left[\int_0^1 \left(2 \int_{-\infty}^{\infty} \frac{\sin \phi}{\phi} d\phi \right)^2 dx \right]^2 (1 + o(1)) \\
&= O\left((b_1 b_2)^2\right).
\end{aligned}$$

Plugging this into (S.27), we have

$$\begin{aligned}
& \int_0^1 \mathbb{K}_b((x_1, x_2, x_3), (r_1, r_2, r_3)) \mathbb{K}_b((x_1, x_2, x_3), (v_1, v_2, v_3)) dx_1 dx_2 dx_3 dr_1 dr_2 dr_3 dv_1 dv_2 dv_3 \\
&= O\left((b_1 b_2 b_3)^2\right) = o(b_1 b_2 b_3).
\end{aligned} \tag{S.28}$$

From (S.23), (S.26) and (S.28),

$$\mu_2 = b_1 b_2 b_3 c_2 + o(b_1 b_2 b_3).$$

The proof is complete. \blacksquare

References

- [1] Andrews, D.W.K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.
- [2] Anselin, L., 2001., Spatial econometrics. A Companion to Theoretical Econometrics. Oxford: Blackwell, 310–330.
- [3] Arellano, M. 1987. Computing robust standard errors for within group estimators. *Oxford Bulletin of Economics and Statistics* 49, 431–434.
- [4] Arellano, M. and Hahn, J. 2005. Understanding bias in nonlinear panel models: some recent developments. In: Blundell, R., Newey, W.K., Persson, T. (Eds.), *Advances in Economics and Econometrics: Theory and Applications*, In: Ninth World Congress, vol. II. Econometric Society Monographs.
- [5] Bertrand, M., Duo, E., and Mullainathan, S., 2004. How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics* 119, 249–275.
- [6] Bester, A., Conley, T., Hansen, C., and Vogelsang, T. 2011. Fixed-b asymptotics for spatially dependent robust nonparametric covariance matrix estimators. Working paper, Michigan State University.
- [7] Bester, C., Conley, T., and Hansen, C. 2011. Inference with dependent data using cluster covariance estimators. *Journal of Econometrics* 165, 137–151.
- [8] Cameron, A., Gelbach, J., and Miller, D. 2011. Robust inference with multi-way clustering. *Journal of Business and Economic Statistics* 28, 238–249.
- [9] Chen, X. and Conley, T., 2001. A new semiparametric spatial model for panel time series. *Journal of Econometrics* 105, 59–83.
- [10] Conley, T., 1999. GMM Estimation with cross sectional dependence. *Journal of Econometrics* 92, 1–45.
- [11] Conley, T. and Dupor, B., 2003. A spatial analysis of sectoral complementarity. *Journal of Political Economy* 111, 311–352.
- [12] Conley, T. and Ligon, E., 2002. Economic distance and cross-country spillovers. *Journal of Economic Growth* 7, 157–187.
- [13] Driscoll, J. and Kraay, A., 1998. Consistent covariance matrix estimation with spatially dependent panel data. *Review of Economics and Statistics* 80, 549–560.
- [14] Fama, E. and MacBeth, J., 1973. Risk, return, and equilibrium: empirical tests. *Journal of Political Economy* 81, 607–636.
- [15] Gonçalves, S., 2011. The moving blocks bootstrap for panel linear regression models with individual fixed effects. *Econometric Theory* 27, 1048–1082
- [16] Hansen, C., 2007. Asymptotic properties of a robust variance matrix estimator for panel data when T is large. *Journal of Econometrics* 141, 597–620.

- [17] Ibragimov, R. and Müller, U., 2010. t-Statistic based correlation and heterogeneity robust inference. *Journal of Business and Economic Statistics* 28, 453–468.
- [18] Jansson, M., 2004. The error in rejection probability of simple autocorrelation robust tests. *Econometrica* 72, 937–946.
- [19] Kelejian, H. and Prucha, I., 2007. HAC Estimation in a spatial framework. *Journal of Econometrics* 140, 131–154.
- [20] Kiefer, N. and Vogelsang, T., 2002a. Heteroskedasticity-autocorrelation robust testing using bandwidth equal to sample size. *Econometric Theory* 18, 1350–1366.
- [21] Kiefer, N. and Vogelsang, T., 2002b. Heteroskedasticity-autocorrelation robust standard errors using the Bartlett kernel without truncation. *Econometrica* 70, 2093–2095.
- [22] Kiefer, N. and Vogelsang, T., 2005. A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory* 21, 1130–1164.
- [23] Kiefer, N., Vogelsang, T., and Bunzel, H., 2000. Simple robust testing of regression hypotheses. *Econometrica* 68, 695–714.
- [24] Kim, M. S. and Sun, Y., 2011. Spatial heteroskedasticity and autocorrelation consistent estimation of covariance matrix. *Journal of Econometrics* 160, 349–371.
- [25] Lin, C. and Sakata, S., 2012. On long-run covariance matrix estimation with the truncated flat kernel. In: Chen, X. and Swanson, N.R. (Eds.), *Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis: Essays in Honour of Halbert L. White Jr*, Springer, New York.
- [26] LeSage, J. and Kelley, P., 2009. *Introduction to spatial econometrics*. CRC Press.
- [27] Mariano, R., 1972. The existence of moments of the ordinary least squares and two-stage least squares estimators. *Econometrica* 40, 643–652.
- [28] Nagar, A. L., 1959. The bias and moment matrix of the general k-class estimators of the parameters in simultaneous equations. *Econometrica* 27, 573–595.
- [29] Newey, W. and West, K., 1987. A Simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.
- [30] Neyman, J. and Scott, E., 1948. Consistent estimates based on partially consistent observations. *Econometrica* 16, 1–32.
- [31] Petersen, M., 2009. Estimating standard errors in finance panel data sets: comparing approaches. *Review of Financial Studies* 22, 435–480.
- [32] Politis, D., 2011. Higher-order accurate, positive semi-definite estimation of large-sample covariance and spectral density matrices. *Econometric Theory* 27, 703–744.
- [33] Sun, Y., 2010. Let’s fix It: fixed-b asymptotics versus small-b asymptotics in heteroscedasticity and autocorrelation robust inference. Working paper, UCSD.

- [34] Sun, Y. and Kaplan, D., 2011. A new asymptotic theory for vector autoregressive long-run variance estimation and autocorrelation robust testing. Working paper, UCSD.
- [35] Sun, Y. and Kim, M. S., 2012. Simple and powerful GMM over-identification tests with accurate size. *Journal of Econometrics* 166, 267–281.
- [36] Sun, Y. and Kim, M. S., 2013. Asymptotic F test in a GMM framework with cross sectional dependence. Working paper, UCSD.
- [37] Sun, Y. and Phillips, P., 2009. Bandwidth choice for interval estimation in GMM regression. Working paper, UCSD.
- [38] Sun, Y., Phillips, P., and Jin, S., 2008. Optimal bandwidth selection in heteroskedasticity-autocorrelation robust testing. *Econometrica* 76, 175–194.
- [39] Vogelsang, T., 2012. Heteroskedasticity, autocorrelation, and spatial correlation robust inference in linear panel models with fixed-effects. *Journal of Econometrics* 166, 303–319.
- [40] White, H. 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48, 817–838.
- [41] Wooldridge, J., 2003. Cluster-sample methods in applied econometrics. *American Economic Review* 93, 133–138.
- [42] Yu, J., de Jong, R., and Lee, L., 2008. Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. *Journal of Econometrics* 146, 118–134.

Table 1: RMSE/Estimand with \hat{J}_{nT} and $\hat{J}_{nT}^{DK} - \text{DGP1}$

λ		θ								
		0.0	0.3	0.6	0.9		0.0	0.3	0.6	0.9
T=15, n=49										
0.0	\hat{J}_{nT} $(\hat{d}_n, \hat{d}_T)_{PA}$	0.10	0.20	0.28	0.46	\hat{J}_{nT} $(d_n^*, d_T^*)_{PA}$	0.42	0.21	0.25	0.43
0.3		0.17	0.34	0.46	0.67		0.14	0.33	0.43	0.60
0.6		0.23	0.43	0.56	0.72		0.17	0.41	0.53	0.71
0.9		0.36	0.55	0.67	0.84		0.31	0.53	0.65	0.83
0.0	\hat{J}_{nT} $(\hat{d}_n, \hat{d}_T)_{RE}$	0.15	0.25	0.32	0.40	\hat{J}_{nT} $(d_n^*, d_T^*)_{RE}$	1.00	0.36	0.36	0.36
0.3		0.24	0.40	0.52	0.71		0.20	0.31	0.41	0.53
0.6		0.32	0.51	0.65	0.69		0.20	0.42	0.52	0.64
0.9		0.41	0.60	0.72	0.84		0.20	0.57	0.63	0.81
0.0	\hat{J}_{nT}^{DK} (\hat{d}_T^{DK})	0.47	0.47	0.48	0.47	\hat{J}_{nT}^{DK} (d_T^{DK})	0.36	0.36	0.36	0.36
0.3		0.55	0.55	0.54	0.55		0.56	0.56	0.55	0.56
0.6		0.69	0.69	0.69	0.69		0.71	0.71	0.71	0.71
0.9		0.88	0.88	0.88	0.88		0.88	0.89	0.89	0.89
T=50, n=49										
0.0	\hat{J}_{nT} $(\hat{d}_n, \hat{d}_T)_{PA}$	0.06	0.13	0.19	0.40	\hat{J}_{nT} $(d_n^*, d_T^*)_{PA}$	0.42	0.12	0.18	0.39
0.3		0.11	0.25	0.34	0.57		0.13	0.24	0.32	0.51
0.6		0.15	0.35	0.50	0.63		0.14	0.31	0.40	0.58
0.9		0.27	0.49	0.59	0.81		0.21	0.44	0.57	0.75
0.0	\hat{J}_{nT} $(\hat{d}_n, \hat{d}_T)_{RE}$	0.08	0.16	0.19	0.21	\hat{J}_{nT} $(d_n^*, d_T^*)_{RE}$	1.00	0.20	0.20	0.20
0.3		0.15	0.28	0.41	0.58		0.20	0.26	0.31	0.33
0.6		0.21	0.44	0.70	0.57		0.20	0.32	0.39	0.47
0.9		0.37	0.61	0.69	0.79		0.20	0.48	0.57	0.69
0.0	\hat{J}_{nT}^{DK} (\hat{d}_T^{DK})	0.28	0.28	0.28	0.28	\hat{J}_{nT}^{DK} (d_T^{DK})	0.20	0.20	0.20	0.20
0.3		0.40	0.40	0.40	0.40		0.38	0.37	0.37	0.38
0.6		0.53	0.54	0.53	0.54		0.51	0.52	0.51	0.52
0.9		0.77	0.77	0.77	0.77		0.77	0.77	0.77	0.77
T=15, n=196										
0.0	\hat{J}_{nT} $(\hat{d}_n, \hat{d}_T)_{PA}$	0.05	0.13	0.19	0.30	\hat{J}_{nT} $(d_n^*, d_T^*)_{PA}$	0.43	0.20	0.21	0.27
0.3		0.09	0.25	0.35	0.55		0.07	0.24	0.32	0.46
0.6		0.14	0.32	0.43	0.57		0.12	0.30	0.39	0.56
0.9		0.30	0.43	0.53	0.71		0.28	0.42	0.51	0.69
0.0	\hat{J}_{nT} $(\hat{d}_n, \hat{d}_T)_{RE}$	0.08	0.17	0.23	0.34	\hat{J}_{nT} $(d_n^*, d_T^*)_{RE}$	1.00	0.36	0.36	0.36
0.3		0.14	0.28	0.44	0.69		0.10	0.26	0.31	0.44
0.6		0.19	0.41	0.57	0.65		0.10	0.31	0.39	0.54
0.9		0.24	0.45	0.58	0.75		0.10	0.36	0.50	0.67
0.0	\hat{J}_{nT}^{DK} (\hat{d}_T^{DK})	0.46	0.47	0.46	0.47	\hat{J}_{nT}^{DK} (d_T^{DK})	0.36	0.36	0.36	0.36
0.3		0.54	0.57	0.55	0.55		0.55	0.56	0.56	0.56
0.6		0.69	0.69	0.69	0.69		0.71	0.71	0.71	0.71
0.9		0.88	0.88	0.88	0.88		0.89	0.89	0.89	0.89

The subscripts ‘PA’ and ‘RE’ denote the Parzen and rectangular kernels, respectively. Left and right panels are based on data-driven bandwidths and infeasible bandwidths, respectively.

Table 2: Empirical Coverage Probabilities of Nominal 95% CIs Constructed
Using Alternative Covariance Estimators - DGP2

λ		θ													
		0.0	0.3	0.6	0.9		0.0	0.3	0.6	0.9		0.0	0.3	0.6	0.9
T=15, n=49															
0.0	PHAC (PA,F)	94.4	93.5	91.7	86.8	DK (PA,F)	95.3	95.2	95.0	93.8	PHAC (RE,F)	94.2	94.1	92.6	91.7
0.3		92.7	91.5	90.1	85.1		93.5	93.3	93.5	92.3		92.7	92.2	91.3	90.1
0.6		89.1	88.8	85.5	79.8		88.5	87.9	88.8	88.0		89.8	89.8	88.7	87.6
0.9		86.5	85.4	82.1	76.1		80.2	81.2	81.0	79.9		86.0	85.1	85.4	80.0
0.0	PHAC _e (PA,F)	94.4	93.5	91.1	86.3	CCE (I)	94.1	92.4	84.5	54.0	PHAC _e (RE,F)	94.2	93.8	92.5	88.7
0.3		92.5	91.2	89.3	84.3		94.0	91.8	84.9	53.5		92.1	91.5	90.5	86.7
0.6		88.6	88.0	84.7	81.0		93.5	91.8	83.9	53.9		88.6	89.1	87.4	84.2
0.9		85.5	84.1	81.1	81.0		93.3	90.8	83.9	54.3		85.0	84.5	83.3	80.4
0.0	PHAC (PA,I)	94.1	93.3	90.7	84.6	DK* (PA,F)	95.0	94.0	91.8	86.1	PHAC (RE,I)	93.9	93.5	91.1	88.2
0.3		92.3	91.2	89.2	82.2		94.7	93.9	91.6	85.0		92.1	91.2	89.8	85.9
0.6		88.4	87.9	83.3	72.8		94.6	93.8	91.7	84.9		88.2	88.2	84.2	74.9
0.9		85.2	83.7	77.9	64.4		94.7	93.2	91.2	85.1		82.6	81.9	75.0	63.0
T=50, n=49															
0.0	PHAC (PA,F)	94.7	94.0	92.8	88.2	DK (PA,F)	95.2	95.2	94.9	94.3	PHAC (RE,F)	94.6	94.5	93.7	93.9
0.3		93.6	93.1	91.9	87.9		94.0	94.1	94.4	93.7		94.0	93.8	93.3	93.5
0.6		92.4	91.5	89.6	84.9		91.3	91.7	91.2	91.1		92.4	92.4	91.6	92.2
0.9		88.9	88.6	85.9	76.0		84.9	85.2	85.1	84.7		88.6	88.8	89.0	85.1
0.0	PHAC _e (PA,F)	94.7	93.8	92.1	87.1	CCE	94.2	92.0	84.7	54.3	PHAC _e (RE,F)	94.5	94.3	93.2	91.2
0.3		93.3	92.6	90.8	86.3		93.9	92.3	84.6	54.6		93.4	93.0	91.2	89.1
0.6		91.9	91.2	88.5	86.4		94.1	92.4	84.4	54.2		92.4	92.0	90.3	89.3
0.9		88.1	87.6	85.9	79.9		93.2	92.4	84.8	53.3		88.1	88.2	87.5	81.7
0.0	PHAC (PA,I)	94.6	93.9	92.5	87.4	DK* (PA,F)	95.5	94.1	92.4	84.8	PHAC (RE,I)	94.5	94.3	93.2	93.0
0.3		93.5	92.9	91.4	86.7		95.0	94.2	92.2	86.6		93.7	93.3	91.9	91.4
0.6		92.1	91.2	87.9	80.7		95.0	94.1	92.1	86.5		91.9	91.7	88.7	83.5
0.9		87.9	87.2	80.5	68.9		93.8	93.6	92.1	85.2		86.4	85.9	77.2	71.8
T=15, n=196															
0.0	PHAC (PA,F)	93.9	93.5	93.3	90.9	DK (PA,F)	94.5	94.2	95.0	94.8	PHAC (RE,F)	93.7	93.7	94.0	92.8
0.3		92.5	93.1	91.7	88.3		92.9	93.9	93.9	92.6		92.5	93.6	92.4	91.3
0.6		91.0	90.2	88.7	86.2		88.7	89.0	88.5	88.6		90.6	90.6	90.3	91.0
0.9		88.3	89.3	87.3	81.9		80.0	81.4	80.4	79.3		87.8	88.3	87.9	83.8
0.0	PHAC _e (PA,F)	94.0	93.1	92.9	90.1	CCE	94.4	93.0	86.9	51.0	PHAC _e (RE,F)	93.9	93.6	93.8	91.8
0.3		92.4	92.5	91.0	87.0		94.6	92.9	86.4	50.6		92.0	92.8	91.9	89.2
0.6		90.2	89.0	87.4	86.2		94.3	93.2	86.0	51.4		90.1	90.0	89.7	88.7
0.9		87.9	88.2	85.8	87.2		94.5	93.2	86.0	50.6		87.0	87.8	87.2	85.6
0.0	PHAC (PA,I)	93.9	93.3	92.9	89.4	DK* (PA,F)	94.2	93.0	94.0	91.2	PHAC (RE,I)	93.6	93.5	93.4	90.5
0.3		92.4	93.0	91.2	86.2		94.6	94.0	93.9	91.2		92.3	93.3	91.5	87.1
0.6		90.8	89.8	87.3	79.7		94.5	94.4	93.3	90.7		90.0	89.9	87.6	78.6
0.9		88.1	88.6	84.8	71.6		94.7	94.3	93.7	91.3		87.0	86.8	81.9	66.7

‘PA’ and ‘RE’ denote the Parzen and rectangular kernels respectively.

‘F’ and ‘I’ denote fixed-smoothing and increasing-smoothing respectively.

The superscript ‘e’ denotes measurement errors.

Table 3: Empirical Coverage Probabilities of Nominal 95% Confidence Regions Constructed with Different Number of Restrictions - DGP2

	λ	g=1				g=3			
		θ				θ			
		0.0	0.3	0.6	0.9	0.0	0.3	0.6	0.9
PHAC (PA,F)	0.0	93.6	92.7	90.9	86.5	92.8	91.2	87.8	78.1
	0.3	91.8	91.1	88.9	84.0	90.4	88.0	84.4	74.0
	0.6	89.2	88.3	85.3	79.0	85.4	82.5	77.5	66.6
	0.9	87.3	86.4	82.2	74.4	80.2	77.1	71.4	58.6
PHAC (PA,I)	0.0	93.3	92.3	90.1	84.1	92.3	90.2	84.9	70.3
	0.3	91.5	90.7	88.0	81.0	89.7	86.7	81.1	63.0
	0.6	88.4	87.6	83.0	71.9	83.4	79.7	70.0	44.5
	0.9	86.1	84.7	77.4	63.2	76.4	71.5	57.3	28.5
PHAC (RE,F)	0.0	93.5	93.2	92.0	90.5	92.7	92.0	90.2	86.8
	0.3	91.8	91.1	89.8	89.2	90.2	88.8	86.8	83.5
	0.6	89.5	89.4	88.6	86.6	86.2	85.0	85.4	87.8
	0.9	86.7	86.1	85.7	80.6	79.3	78.0	81.7	84.1
PHAC (RE,I)	0.0	93.1	92.7	90.9	86.6	91.7	90.2	85.9	74.3
	0.3	91.2	90.5	88.3	84.0	88.4	86.5	81.5	66.3
	0.6	88.0	87.9	84.6	73.0	81.0	78.5	69.4	51.2
	0.9	83.5	82.1	75.1	62.5	68.0	64.3	51.6	53.4

See notes to Table 2.

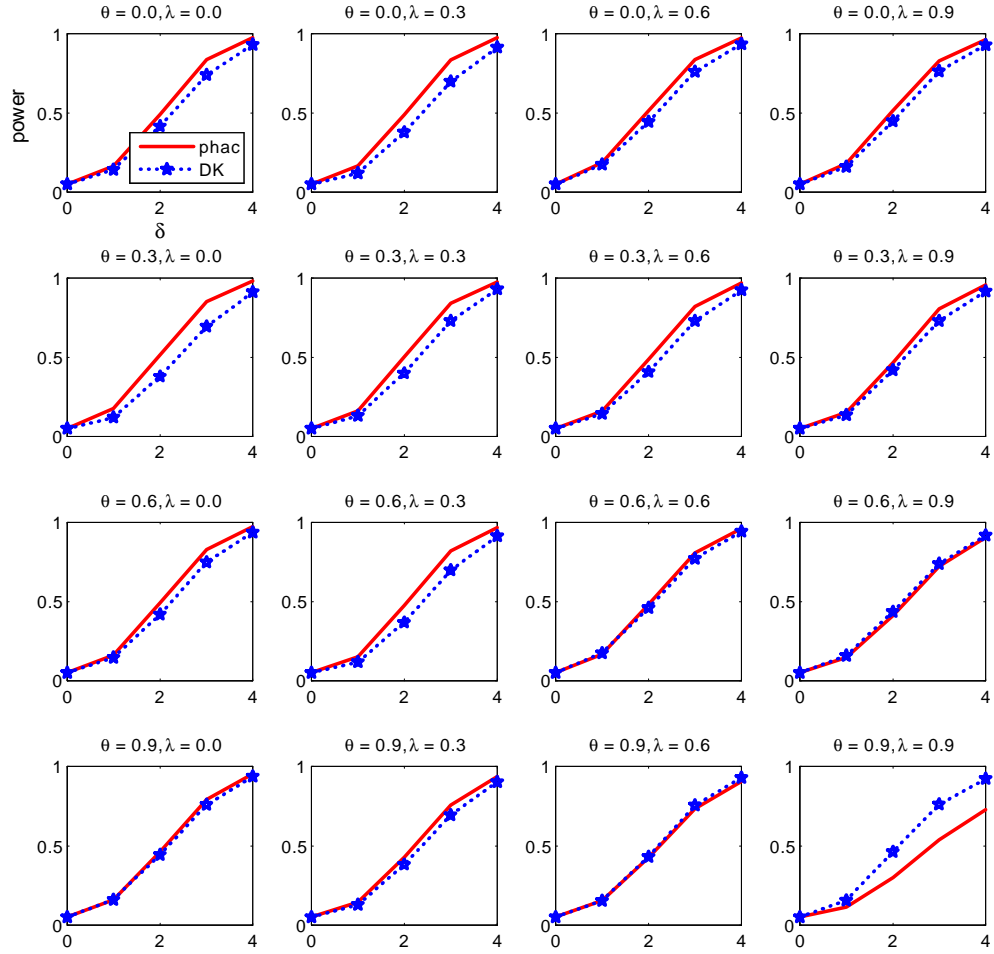


Figure 3: Size-adjusted power of the PHAC and DK with $n = 49, T = 15$.